Cloning of citrate synthase from the halophilic archaeon Haloferax volcanii

James, Keith D.

Award date:
1994

Awarding institution:
University of Bath

Cloning of Citrate Synthase from the Halophilic Archaeon *Haloferax volcanii*

submitted by Keith D. James

for the degree of PhD

of the University of Bath

1994

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The effect of monovalent salt on the activity and stability of citrate synthase from the moderately halophilic Archaeon *Haloferax volcanii* was investigated by assay of enzyme activity over a range of Na^+ and K^+ concentrations and by thermal inactivation experiments over a range of K^+ concentrations.

Based on a 38 residue N-terminal amino acid sequence obtained. from *Hf.volcanii* citrate synthase, oligonucleotides were designed and used to probe restriction digests of *Hf.volcanii* genomic DNA. A 2.9 kb genomic fragment was identified by hybridisation to one of these probes. This fragment did not contain the citrate synthase gene; however, a 939 bp open reading frame within this fragment showed significant (33-47%) identity with several bacterial genes involved in the metabolism of sugar-nucleotides. The structural features of this open reading frame are discussed and possible functions in glycolipid and glycoprotein biosynthesis suggested.

To aid the cloning of the citrate synthase, further amino acid sequence data was obtained from the protein. Citrate synthase was purified by an improved affinity chromatography method, using biospecific elution. Cyanogen bromide cleavage of the purified enzyme yielded a 6 kD fragment. Amino acid sequence internal to the enzyme was obtained from this fragment.

PCR primers were designed to the N-terminal and internal amino acid sequences and used to amplify a 960 bp product from genomic DNA. This product is of the size predicted from the relative positions of the amino acid sequences used to design the PCR primers. Due to time constraints the product remains to be cloned and sequenced.

Table of Contents

List of Figures

Abbreviations

A_{280}	absorbance at 280 nm
BSA	bovine serum albumin
CoA	coenzyme A
CS	citrate synthase (EC 4.1.3.7)
dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytidine 5'-triphosphate
DEAE	diethylaminoethyl
dGTP	deoxyguanosine 5'-triphosphate
DNA	deoxyribonucleic acid
DTNB	5,5'-dithiobis-(2-nitrobenzoic acid)
DTT	dithiothreitol
dTTP	deoxythymidine 5'-triphosphate
EDTA	(disodium) ethylenediamine tetraacetate
IPTG	isopropyl-β-D-thiogalactoside
kb	kilobase pairs
kD	kilodaltons
LMP Agarose	low melting point agarose
MDH	malate dehydrogenase (EC 3.1.1.1.37)
M_r	relative molecular mass
ORF	open reading frame
RNAse	ribonuclease A (EC 3.1.27.5)
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
TAE	tris-acetate EDTA
TBE	tris-borate EDTA
TCA	trichloroacetic acid

TEMED	N,N,N',N'-tetramethylethylene diamine
T_m	Melting temperature
Tris	tris-(hydroxymethyl)-methylamine
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galctoside

Acknowledgements

I would like to thank Dr. David Hough and Dr.Mike Danson for advice, patience and many helpful discussions. Thanks to Janice Young of Zeneca for cyanogen bromide cleavage and N-terminal sequencing and Dr. David Byrom of ICI for his advice.

Finally, thanks to all the people of the "end" and "middle" labs for their help and for generally making my stay here enjoyable.

I would like to acknowledge that the initial purification was carried out by Dr. Maria Bonete of the University of Alicante and that the citrate synthase multiple sequence alignment was carried out by Michelle McCormack.

Financial support for this work was provided by the SERC and by ICI, through a CASE award.

Chapter 1

Introduction

1.1 The diversity of Archaeal phenotypes

A striking feature of the group of organisms that we know as the archaebacteria, or Archaea, is their range of extreme requirements and tolerances for growth. Although in general their phenotypes can be thought of as either methanogenic, thermophilic or halophilic, within these broad categories are found autotrophic, heterotrophic, anaerobic, aerobic, acidophilic and alkaliphilic representatives.

The methanogens are obligate anaerobes, some of which are thermophilic or halophilic, that metabolise carbon dioxide to methane. Thermophiles may have growth optima anywhere between 55 and 110°C, some also requiring highly acidic conditions. A large proportion of them reduce sulphate or elemental sulphur to hydrogen sulphide to obtain energy. The halophilic Archaea require salt concentrations of greater than 1 M NaCl for growth and include representatives that also require highly alkaline conditions; their internal salt environments are isotonic with the exterior, but KCl replaces NaCl.

In terms of cellular and sub-cellular morphology there is little to distinguish the Archaea from the eubacteria, a fact which originally led to their taxonomic grouping with the latter as prokaryotes. Their physiological peculiarities, such as ether-linked lipids and unusual cell wall structure were

ascribed to convergent evolution and their extremophilic natures to adaptations to their environments. Molecular analysis has since revealed that the Archaea must be considered as phylogenetically distinct from both the eubacteria and the eukaryotes.

1.2 Archaeal Phylogeny

The way in which we view the evolutionary history of life was revolutionised by the first thorough analyses of 16s rRNA sequence data and subsequent inference of previously unknown phylogenetic relationships (Woese & Fox, 1977, Woese *et al.*, 1978). The prokaryote/eukaryote dichotomy was replaced, after some controversy, by the division of life into eukaryotes, eubacteria and archaebacteria that is now generally accepted. It has recently been proposed that these groups be given the status of "domains" (Eukarya, Bacteria and Archaea respectively), the archaebacteria being renamed the Archaea to emphasise their distinction from the eubacteria (Woese *et al.*, 1990).

The Archaea are sub-divided into two main branches: the Crenarchaeota, comprising mainly acidophilic and sulphur-dependent thermophiles and the Euryarchaeota, which include the extreme halophiles, the methanogens, sulphate-reducers and some thermophiles (Figure 1.1).

1.2.1 The root of the universal tree

The position of the root shown on the 16S rRNA tree (Woese *et al.*, 1990) has been determined from the data of Gogarten *et al.* (1989) and Iwabe *et al.* (1989) who used pairs of paralogous genes (that were assumed to have arisen by duplication events in the last common ancestor of all life) to locate the root between the Bacteria and Archaea. The root has simply been drawn on the rRNA tree at the equivalent position.

Studies on glutamate dehydrogenase genes (Benachenhou *et al.*, 1993, Forterre *et al.*, 1993) have revealed that they are also suitable for rooting the universal tree by a similar method to that used by Iwabe *et al.*, where one of a pair of ancestrally duplicated genes is used as an outgroup to root the tree of the other. Using several different phylogenetic inference methods these workers did not find it possible to root the tree unambiguously. Forterre *et al.* (1993) have also suggested that the way in which Iwabe *et al.* performed their calculations was not statistically sound. Regardless of the method of analysis, the currently accepted rooting is based on only two gene pairs and therefore needs further testing before it can be regarded as certain.

1.2.2 How valid is the Archaeal tree ?

Since the original proposal for the existence of three branches of life, which was based on partial rRNA sequences, the data have been augmented by complete sequencing of many 16S rRNAs, increasing the resolution of the tree. Phylogenetic trees deduced from 23S rRNA (Leffers *et al.*, 1987), ribosomal proteins (Auer *et al.*, 1990), DNA dependent RNA polymerases

(Zillig *et al.*, 1989a,b), protein elongation factors and ATPases (Iwabe *et al.*, 1989) support the three domain scenario. Acceptance of this relationship has by no means been unanimous; a number of objections have been raised, either to the topology of the 16S rRNA tree (specifically to the monophyletic nature of the Archaea) or to the validity of any tree based on a single molecular feature such as 16S rRNA sequence.

1.2.3 Objections to the Archaeal tree

In the past Lake has argued that the algorithms which support the archaeal tree are subject to errors caused by unequal rate effects, where rapidly changing sequences are artefactually grouped together (Lake, 1988, 1989). He put forward the evolutionary parsimony algorithm as a more suitable alternative on the basis that it is insensitive to unequal rate effects. The branching pattern obtained is different from the archaeal tree: the sulphur-dependent thermoacidophiles group with the eukaryotes while the halophiles and methanogens group with the eubacteria, thus the Archaea appear polyphyletic (Figure 1.2). Although this view has had some support, it has been refuted by the proponents of the archaeal topology (Olsen & Woese, 1989, 1993) on the basis of the regions of sequence chosen for the analysis and the way in which they were aligned.

Zillig *et al.*(1989a,b) do not contest the existence of the Archaea, but have interpreted sequence data from DNA-dependent RNA

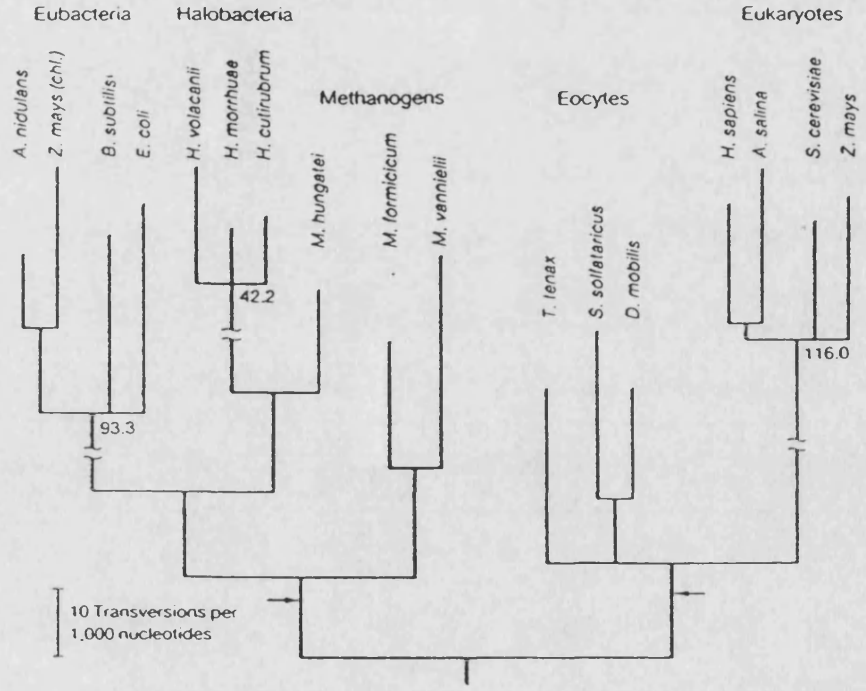

Figure 1.2 Universal phylogenetic tree according to Lake

polymerases as showing that the eukaryotes have arisen from a fusion of an ancestral eubacterium and Archaeon.

More fundamental objections are raised by Margulis and Guerrero (1991) and by Mayr (1990), who dispute the importance of the molecular data and who see the structural and genetic differences between "prokaryotes" and eukaryotes as paramount. For these reasons Margulis and Guerrero would prefer a five kingdom system to be adopted and Mayr a system based on two domains. Their proposals are discussed more fully in Mayr (1991), Woese *et al.* (1991) and Wheelis *et al.* (1992).

1.2.4 A conflict of assumptions, not of data

The arguments surrounding the validity of the various trees stem mainly from differences of opinion on the aims of classification (to show evolutionary relationships, or to act as an information retrieval system) and on the mathematical treatment of the available data. There are several stages at which critical choices must be made: truly homologous sequences must be chosen and aligned with minimal gaps, regions of the alignment suitable for analysis must be identified and finally the phylogenetic inference method(s) selected. There may be a number of equally valid options at each stage. (For a review see Beanland & Howe, 1992.) From the available data the division of life into three domains appears safe, although the current rooting of the tree must be taken with some caution.

1.3 *Haloferax volcanii*

Haloferax volcanii is a moderately halophilic Archaeon originally isolated from the Dead Sea (Mullakhanbhai & Larsen, 1975). At 30°C it requires a minimum of 1 M NaCl for growth, the optimum being 1.7 M NaCl. *Hf.volcanii* has a high tolerance for magnesium, as would be expected for an organism isolated from the Dead Sea, which contains 1.1 - 1.5 M Mg^{2+}. It is capable of growing on a simple, defined minimal medium, allowing the isolation of auxotrophic strains (Mevarech & Werczberger, 1985).

The *Hf.volcanii* genome consists of a circular 2.9 Mbp chromosome and four plasmids of 690, 442, 89 and 6.4 kbp. Detailed physical mapping of the *Hf.volcanii* genome has yielded a set of overlapping cosmid clones that cover 96% of the total sequence (Charlebois *et al.*, 1990, 1991). The *Hf.volcanii* genome contains fewer mobile insertion sequences than those of other halophilic Archaea, (*Halobacterium halobium*, for example) making it relatively stable.

The development of transformation techniques (Charlebois *et al.*, 1987) and *E.coli / Hf.volcanii* plasmid shuttle vectors (Lam & Doolittle, 1989, Holmes *et al.*, 1991) has established *Hf.volcanii* in the field of halophile molecular biology.

1.4 The citric acid cycle

The citric acid cycle is the site of the terminal reactions of aerobic catabolism in eukaryotes and eubacteria, acting as a source of reducing equivalents for ATP synthesis (Figure 1.3). The cycle also provides a pool of biosynthetic precursors for anabolic processes.

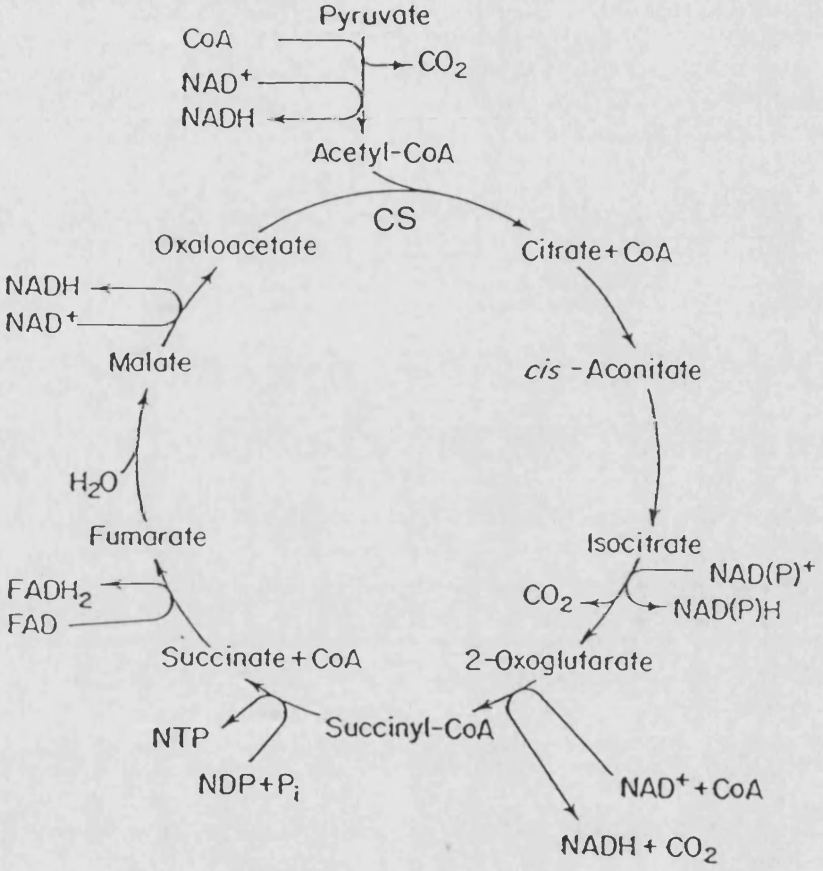

Figure 1.3 The citric acid cycle

The position of citrate synthase is indicated (CS).

There has been no thorough study on the presence of the oxidative cycle in Archaea. All of the cycle's enzymes have been reported in *Halobacterium halobium* (Aitken & Brown, 1969) and some in aerobically grown *Thermoplasma acidophilum* (Danson, 1988). Key enzymes (citrate synthase and succinate thiokinase) have been reported in a number of halophiles (Danson *et al.*, 1985). It is likely that a complete cycle is operative in these organisms.

An oxidative cycle may operate in sulphur-dependent Archaea growing heterotrophically and aerobically (*Sulfolubus* species, for example) as some of the required enzymes activities have been observed [see Danson, (1988) for review]. Those sulphur-dependent Archaea that grow autotrophically also possess the majority of the citric acid cycle enzymes, however they appear to be involved in CO_2 fixation via a reductive pathway (Schäfer *et al.*, 1986)

1.5 Citrate synthase

1.5.1 The available data

Citrate synthase (EC 4.1.3.7) acts as a major control point in the citric acid cycle, catalysing the condensation of oxaloacetate and acetyl coenzyme A to form citrate and coenzyme A. The enzymes of the Gram negative bacteria are hexameric and are allosterically inhibited by NADH, while those of Gram positive bacteria, eukaryotes and Archaea are dimeric and are insensitive to NADH, but are isosterically inhibited by ATP. In all cases the subunit relative molecular mass is 45-50 kD.

As a consequence of its metabolic importance, citrate synthase has been extensively studied. The gene sequences of number of citrate

synthases have been determined from a wide variety of sources, shown in Table 1.1

pig kidney	(Evans *et al.*, 1988)
Saccharomyces cerevisiae mitochondria	(Rosenkrantz *et al.*, 1986)
Saccharomyces cerevisiae glyoxysome	(Rosenkrantz *et al.*, 1986)
Arabidopsis thaliana	(Unger *et al.*, 1989)
Escherichia coli	(Ner *et al.*,1983)
Acinetobacter anitratum	(Donald *et al.*, 1987)
Acetobacter aceti	(Fukaya *et al.*, 1990)
Pseudomonas aeruginosa	(Donald *et al.*, 1989)
Bacillus coagulans	(Schendel *et al.*, 1992)
Rickettsia prowazekii	(Wood *et al.*, 1987)
Coxiella burnetii	(Heinzen *et al.*, 1991)
Tetrahymena thermophila	(Numata *et al.*, 1991)
Mycobacterium smegmatis	(David *et al.*, 1991)
Thermoplasma acidophilum	(Sutherland *et al.*, 1990)

Table 1.1 Citrate synthase gene sequences currently available

Crystal structures are available for both the pig and chicken heart enzymes (Remington *et al.*, 1982, Liao *et al.*, 1991). Preliminary crystallographic data have recently been obtained for the citrate synthase of the Archaeon *Tp.acidophilum* (Russell *et al.*, 1993).

The crystallographic data show that the enzyme is composed almost entirely of α-helices. Each subunit consists of a large and small domain, with the active site located in a cleft between them. Each subunit contributes residues to the active site of the other. Multiple sequence alignments (Henneke *et al.*, 1990) have shown that the majority of the twelve active site residues are conserved between all eubacterial and eukaryotic citrate synthases.

Two forms of the enzyme have been observed: the "open" form in which the active site is accessible to the solvent and the "closed" form where the solvent has no access to the active site, providing acetyl coenzyme A is bound. The open form has a binding site for oxaloacetate in the active site, but none for acetyl coenzyme A. When oxaloacetate binds, the large and small domains rotate relative to each other and the enzyme assumes the closed form, which can then bind acetyl coenzyme A. The open form mediates substrate entry and product release, while the closed form carries out catalysis (Remington, 1992). Figure 1.4 shows a diagram the secondary structure of a pig heart citrate synthase monomer and space-filled representations of its dimer in both open and closed forms.

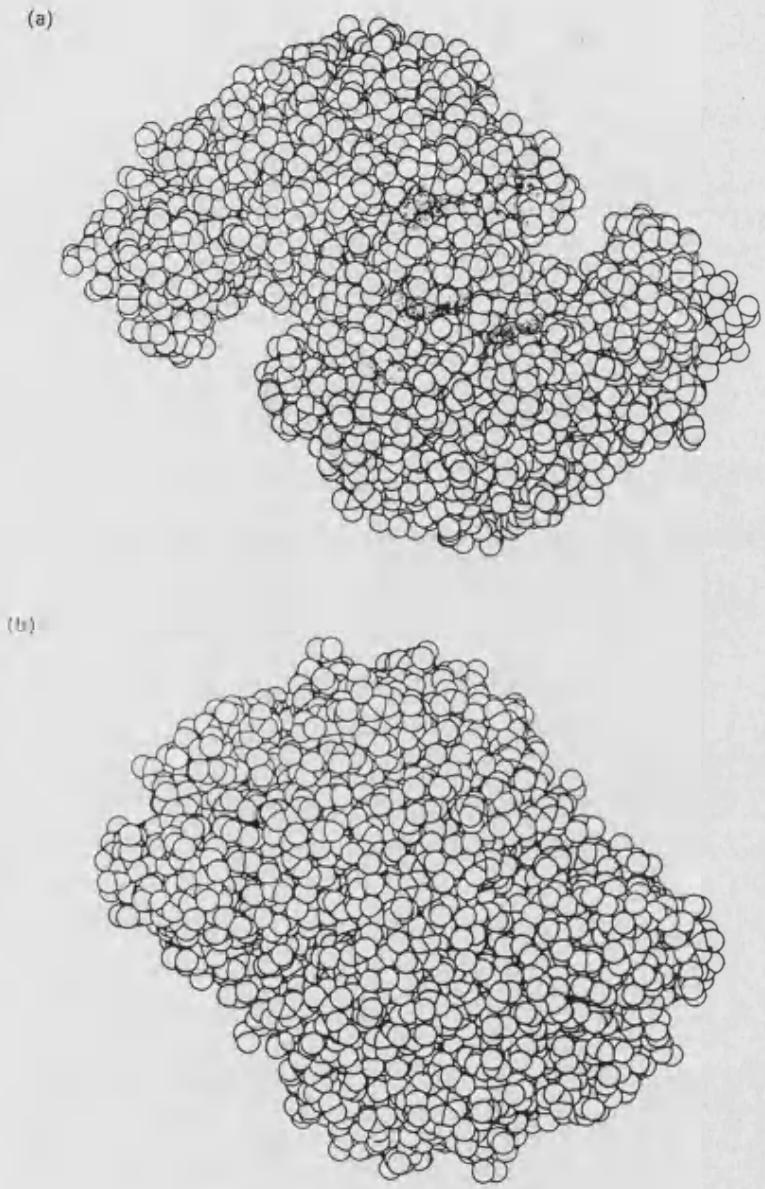

Figure 1.4 Structural diagrams of pig heart citrate synthase

"Open" (a) and "closed" (b) forms of the pig heart citrate synthase dimer, from Remington (1992). The two obvious clefts in the open form indicate the position of the active sites.

1.5.2 The citrate synthase mechanism

In order to bring about the condensation of acetyl coenzyme A and oxaloacetate, citrate synthase must first deprotonate the methyl carbon of the acetate moiety, using only the weak acids and bases of its amino acid side chains. Simultaneously, oxaloacetate must be activated, making it a suitable target for nucleophilic attack.

The catalytic mechanism has been extensively studied in pig heart citrate synthase (Karpusas *et al.*, 1990, Alter *et al.*, 1990). From studies of the binding of acetyl coenzyme A and the proposed transition state analogue carboxymethyl coenzyme A, residues His274 and Asp375 have been identified as acting in concert on acetyl coenzyme A (Karpusas *et al.*, 1990, Karpusas *et al.*, 1991). These residues are thought to promote formation of a neutral enol by general acid-base catalysis. Almost any mutation of the catalytic residues results in a substantial increase of the enzyme's stability to thermal inactivation (Wang *et al.*, 1991). This was taken to indicate that in the native enzyme these residues are held in unfavourable positions such that their pK_a values are radically changed from their solution values, enabling them to deprotonate the methyl carbon.

Bound oxaloacetate is surrounded by three arginines and two histidines, polarizing the carbonyl bond and making it prone to nucleophilic attack (Karpusas *et al.*, 1990). The stereospecificity of the reaction is explained by the binding of oxaloacetate, as only one side of the molecule is open to attack, the other being obscured by the enzyme.

The subsequent condensation step is proposed to occur as the planar

enol attacks the C3 of oxaloacetate and the proton originally donated by His274 is recycled by passing back to that residue. Figure 1.5a summarises these steps.

The resulting citryl coenzyme A intermediate is then hydrolysed by a mechanism that has not been fully characterised. Remington (1992) has suggested that hydrolysis by an activated water molecule may occur (Figure 1.5b), while Alter *et al.* (1990) suggest that an anhydride is formed with Asp375 prior to a similar hydrolysis by an activated water molecule. After the condensation step Asp375 is protonated, enabling it to donate the proton to the thioester sulphur atom and thereby activate it, since sulphonium ions are good leaving groups (Figure 1.5c).

Mutation of the active site aspartate of the *Escherichia coli* citrate synthase (Man *et al.*, 1991) has shown that this residue is involved in the hydrolysis step. When changed to glutamate the forward reaction rate is greatly reduced, although the hydrolysis of citryl-coenzyme A is unaffected. This has led them to propose a reaction scheme which does not involve the aspartate residue until the hydrolysis step, assuming that a carbanion is formed, instead of a neutral enol (Figure 1.5d). They propose that the developing carbanion is stabilised by the partial positive charge on the carbonyl of oxaloacetate. The *E.coli* enzyme may therefore operate by a slightly different mechanism to the eukaryotic enzyme.

1.5.3 Citrate synthase as a model enzyme

The extensive database of structural and mechanistic information available for citrate synthase make it an ideal candidate for comparative

molecular enzymology. It is a widely occurring enzyme which has a conserved function in all organisms where it occurs and as such is suitable for phylogenetic analysis.

Citrate synthase has been reported in the three major groups of Archaea (Danson *et al.*, 1985), presenting an opportunity to compare the structures of thermophilic and halophilic enzymes with those already determined for a variety of mesophiles.

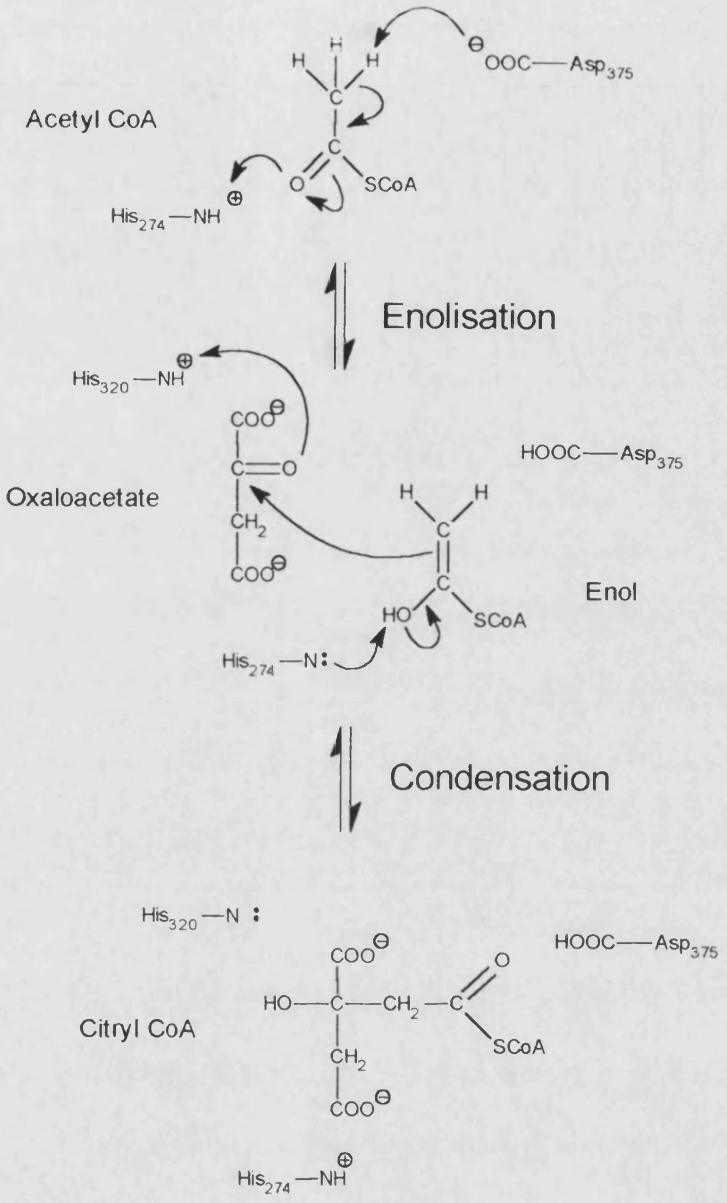

Figure 1.5a Enolisation and condensation steps

This proposed mechanism is based on data from the pig heart enzyme.

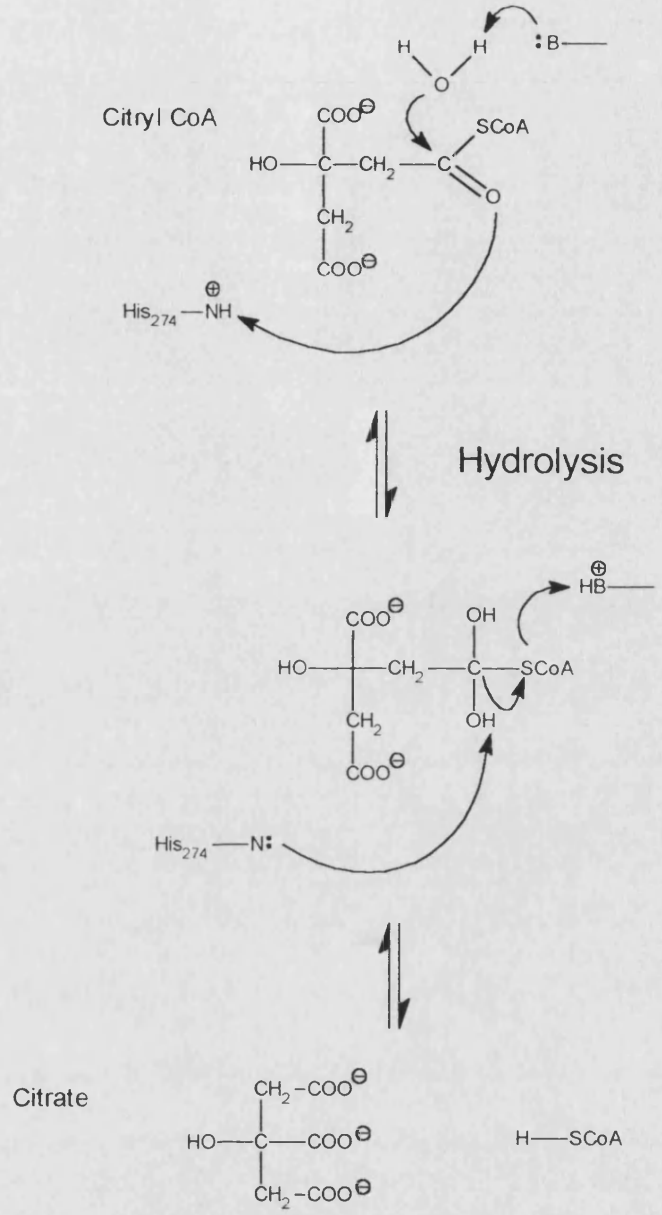

Figure 1.5b Hydrolysis step mechanism

This proposed mechanism is based on data from the pig heart enzyme. B- represents an unknown base on the enzyme surface.

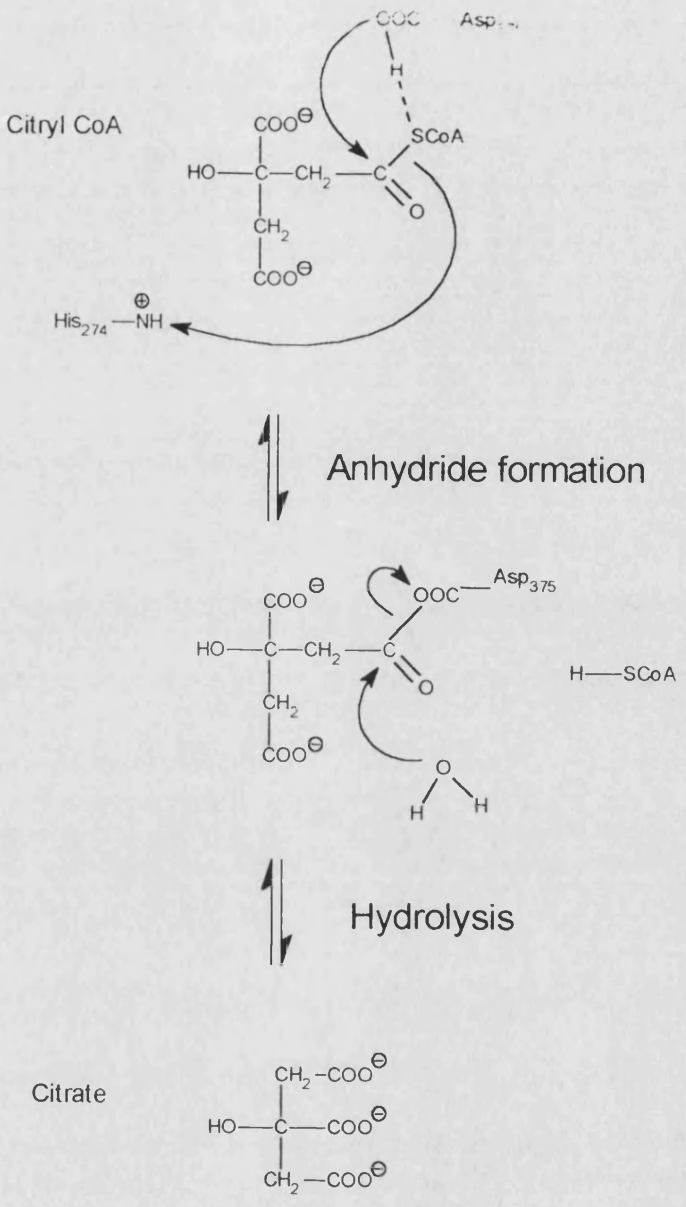

Figure 1.5c Hydrolysis via anhydride formation

This proposed mechanism is based on data from the pig heart enzyme.

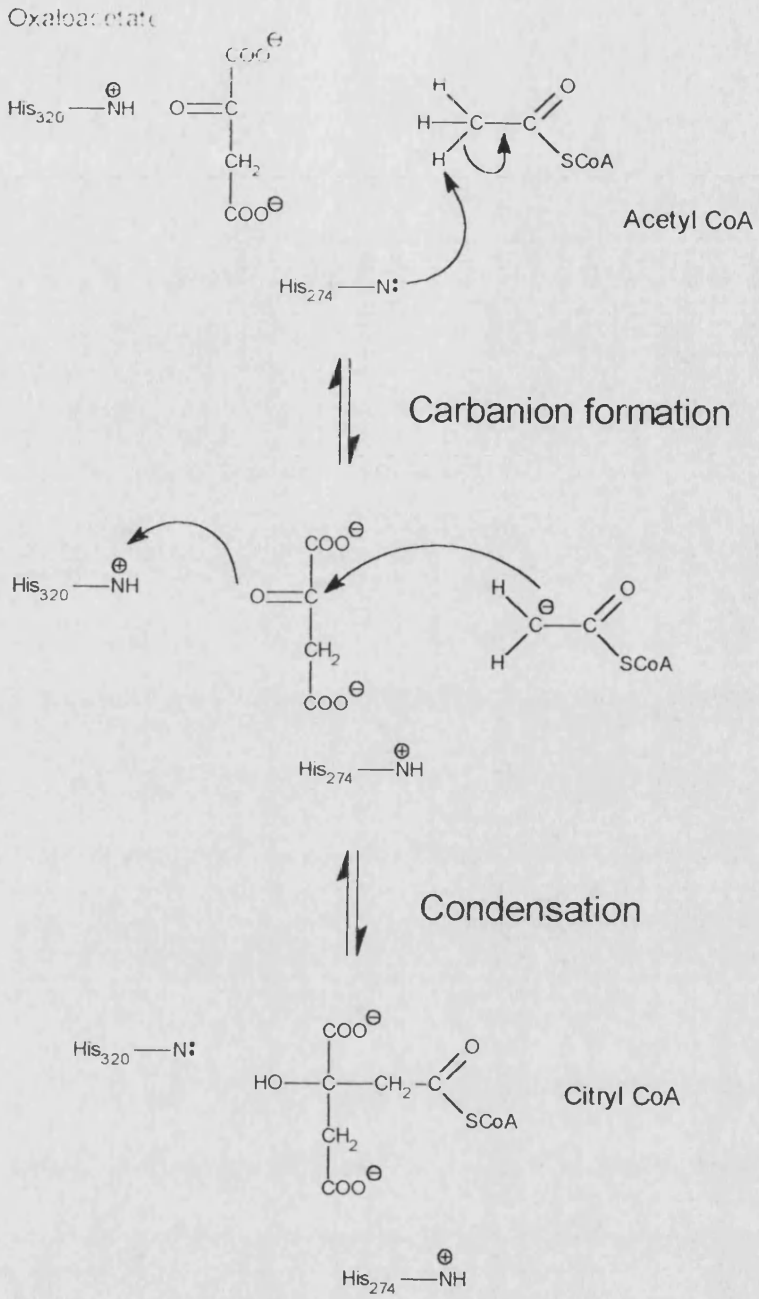

Figure 1.5d Carbanion formation and condensation steps

This proposed mechanism is based on data from the *E.coli* enzyme.

1.6 The structure of halophilic enzymes

The structural features of halophilic enzymes that are responsible for their extreme salt requirements are not well defined. In the absence of any X-ray crystallographic data, techniques such as light, X-ray and neutron scattering and analytical centrifugation have been used to probe enzyme structures in solution.

It is evident from the primary amino acid compostions of halophilic enzymes that they are generally acidic compared to their mesophilic counterparts and contain a higher proportion of borderline hydrophobic residues (Lanyi, 1974). Current data suggest that the acidic residues are clustered in surface loops surrounding a core similar to a mesophilic protein. Lanyi concluded that at low salt concentrations (< 0.5 M) the predominant reason for salt dependence is the shielding of the many negative charges on the protein, while at higher concentrations hydrophobic interactions become more important.

Studies on the malate dehydrogenase (hMDH) of *Haloarcula marismortui* led to the proposal of a model for the stabilization of halophilic proteins under high salt conditions (Zaccai & Cendrin, 1989, Zaccai & Eisenberg, 1990, Eisenberg *et al.*, 1992). It is argued that the acidic groups on the surface of the protein form a quaternary structure that coordinates hydrated salt ions at a higher concentration than in the bulk solvent. The surface carboxyl groups are assumed to act cooperatively as there are otherwise too few to account for the amount of hydrated salt ions bound (approximately 500 Na^+ ions, 500 Cl^- ions and 4000 water molecules per hMDH dimer). This salt binding corresponds to 0.35 g NaCl and 0.85 g water

per gram of protein, an exceptionally high value when one considers that non-halophilic proteins bind 0.2-0.4 g of water and negligible amounts of salt. When dissociated into monomers by lowering the salt concentration, hMDH lost its unusual salt and water binding capacity, behaving more like a non-halophilic protein. This was taken as evidence to support the involvement of quaternary structure in salt binding.

The stabilization model has since been revised following the cloning, sequencing and expression in *E.coli* of the hMDH gene (Cendrin *et al.*, 1993). The original model assumed the enzyme to be dimeric and of subunit relative molecular mass 87 kD, as determined by light and neutron scattering and analytical centrifugation (Zaccai & Cendrin, 1989). The relative molecular mass deduced from the sequence data is however 32.6 kD. The accuracy of data obtained from analytical centrifugation and light scattering has been improved and the neutron and X-ray data re-analysed, leading to the conclusion that hMDH is in fact a tetramer of about 140 kD relative molecular mass (Bonnete *et al.*, 1993).

The calculations of the amounts of salt and water bound are significantly affected by these new data; the tetramer appears to bind approximately 0.4 g water and 0.1 g NaCl or KCl per gram of protein. This is about the same amount of water and about ten-fold the amount of salt as a non-halophilic protein. The water binding ability of hMDH does not change significantly upon dissociation of the tetramer in low salt conditions (1 M NaCl); the monomers appear to remain partially folded with the same solvent binding characteristics as the tetramer.

Studies on the elongation factor Tu of *H.marismortui* (hEF-Tu) provide evidence to support this stabilization model (Baldacci *et al.*, 1990). The

translated gene sequence of hEF-Tu was compared with those of *E.coli* and *Methanococcus vanielii* (a methanogenic Archaeon), revealing the expected excess of acidic residues in the hEF-Tu sequence. The predicted antigenicity of hEF-Tu was determined by the method of Parker *et al.* (1986) and compared with the predictions for *E.coli* and *M.vanielii*. The data suggest that nearly all the acidic residues in hEF-Tu are located on the protein surface. By locating the equivalent residues on the X-ray structure of the *E.coli* EF-Tu it appears that the acidic residues in hEF-Tu occur in patches on the exterior of the protein.

Further evidence for the presence of surface charge clusters as a common feature of halophilic proteins has been presented by Bohm and Jaenicke (1994), who constructed a homology model of the *Hf.volcanii* dihydrofolate reductase (hDHFR) based on the crystal structure of the *E.coli* enzyme. These workers considered halophilicity in terms of the enzyme's instability at low salt concentrations, rather than it's ability to function at high salt concentrations.

They observed that hDHFR retained activity under acid conditions, in the absence of monovalent salt. This was explained by suggesting that the driving force behind the inactivation in low salt was the charge repulsion between the negative surface residues no longer shielded from each other by salt ions. Under acid conditions these residues are protonated and electrostatically neutral, preventing disruption of the protein structure.

The hDHFR model identified five surface clusters of 2-3 amino acids each (aspartate and glutamate) that may be relevant to the protein's halophilic properties. One of these clusters occurs at the C-terminal end of an α-helix, an unusual feature as this is the negative end of the helix dipole

(the dipole moment of an α-helix is approximately 0.5 - 0.75 of a negative charge at the C-terminal end, and a similar positive charge at the N-terminal end). Bohm and Jaenicke termed this a "halophilic helix motif", as in general positive charges are preferred at the C-termini of helices (Richardson & Richardson, 1988).

1.7 Biotechnology of halophilic systems

The characteristics of the halophilic Archaea have not yet been widely exploited in biotechnological processes. Present areas of interest include applications of isolated halobacterial purple membrane and the production of biopolymers by *Haloferax mediterranei.*

The photosensitive, proton-pumping membranes of some halobacteria have been isolated in a highly stable and active form which has potential uses in photoelectric cells and photosensitive films (Rodriguez-Valera, 1991).

Under certain growth conditions *Hf.mediterranei* can be made to accumulate the high molecular weight polymer, poly-β-hydroxybutyrate in intracellular granules. This material is used in the manufacture of biodegradable plastics (Rodriguez-Valera *et al.*, 1989).

The unusual water-retaining properties of halophilic enzymes have yet to be applied in any industrial fermentations. These enzymes may find future applications as biocatalysts in non-aqueous media, particularly hydrophilic solvents which are thought to denature enzymes by stripping essential water molecules from them (Klibanov, 1989). Under these very low water conditions enzymes become rigid and as a consequence exhibit novel substrate specificities and increased thermostability (Klibanov,1989, Gupta,

1992). An extracellular protease from *Halobacterium halobium* has been shown to be stable and active in the polar organic solvents DMSO and ethyleneglycol; furthermore, its activity changes from amide hydrolysis to an esterase activity, enabling it to catalyse peptide syntheses (Ryu & Dordick, 1993).

Halophilic enzymes may also find applications in aqueous environments maintained at high solute concentrations, such as in industrial processes where operation at a high substrate or product concentration would be advantageous.

In the long term it may be possible to engineer features that confer halophilicity into enzymes with useful catalytic activities, if such general features can be identified.

1.8 Aims of this project

Using citrate synthase as a model, our group hopes to identify features that confer on archaeal enzymes their stability to extremes of temperature and salinity. This is to be achieved by comparison of high resolution three dimensional structures, coupled with mutational studies to test the deductions.

We aim to clone, sequence and express at high level citrate synthase from a variety of Archaea whose growth optima span the temperature range of 30 to 110°C and salinity range up to saturation. Once sufficient pure protein is available, crystallisation and rational mutagenesis studies will follow.

The protein sequence data inferred from the citrate synthase gene sequences will be used to construct phylogenetic trees, with the aim of

further clarifying the phylogeny of the Archaea and their evolutionary relationship to other organisms.

The halophilic Archaeal citrate synthases we intend to study are those from the moderate halophile, *Haloferax volcanii* and the extreme halophile, *Halobacterium halobium*. As part of our overall scheme, this project aims to clone, sequence and express at high level the gene for citrate synthase of *Hf.volcanii*, either in a mesophilic host, or in a *Hf.volcanii* citrate synthase-deficient mutant. Sequence data obtained will be subject to phylogenetic analysis, while expressed protein will be used to begin crystallisation studies.

Chapter 2

Materials and Methods

2.0 Materials

Chromatography: DEAE-Sepharose CL-6B was supplied by Pharmacia Biotech Ltd., Milton Keynes, UK, Hydroxylapatite Bio Gel HT by Bio-Rad Laboratories Ltd. Herts., UK and Matrex gel Red A by Amicon Corporation, MA, USA.

Citrate synthase assays: Coenzyme A (CoA) was supplied by Calbiochem-Novobiochem, Nottingham, UK, 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) and oxaloacetic acid by Sigma Chemical Co. Ltd., Poole, UK.

Electrophoresis: Acrylamide Sequagel and Protogel were supplied by National Diagnostics Ltd., NJ, USA, and N,N,N'N'-tetramethylethylene diamine (TEMED) by Sigma. Protein molecular weight standards were supplied by Bio-Rad Laboratories Ltd. Agarose (standard and low melting point (LMP)) and ethidium bromide solution were supplied by Sigma.

Blotting: Genescreen transfer membrane was supplied by DuPont NEN Research Products, Stevenage, UK, Hybond N membrane by Amersham International, Aylesbury, UK and Immobilon P transfer membrane by Millipore Ltd., Watford, UK. Whatman 3MM chromatography paper was supplied by Whatman Scientific Ltd., Maidstone, UK.

Bacterial culture: Yeast extract, tryptone, casamino acids and Bacto-agar were supplied by Difco Laboratories, MI, USA. Ampicillin (sodium salt) was supplied by Sigma.

Molecular biology: Restriction enzymes were supplied by Pharmacia and by Gibco BRL Life Technologies Ltd., Uxbridge, UK. Alkaline phosphatase was supplied by Pharmacia, Klenow DNA polymerase and Taq DNA polymerase by Gibco BRL or by Northumbria Biologicals Ltd (NBL), Cramlington, UK. Lambda DNA, Lambda DNA *Hind*III and *Pst*I digests, 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) and isopropyl-β-D-thiogalactoside (IPTG) were supplied by NBL. NENsorb 20 columns were supplied by DuPont NEN and Geneclean II kits by Bio 101, La Jolla, CA, USA. Magic Miniprep kits were obtained from Promega Corporation, WI, USA. Low molecular DNA markers were supplied by Bio-Rad Laboratories Ltd. Sequenase kits were supplied by United States Biochemical Corp., Ohio, USA. Dithiothreitol was supplied by ICN Biomedicals Ltd., High Wycombe, UK. Redistilled phenol was supplied by Rathburn Chemicals Ltd., Walkerburn, UK.

Radiochemicals: Adenosine-5'-triphosphate, tetra (trimethylammonium) salt, [γ ^{32}P] 3000Ci/mmol was supplied by DuPont NEN and deoxyadenosine-5'-(α-thio)-triphosphate, [α^{35}S] 1000Ci/mmol by ICN Biomedicals Ltd.

Oligonucleotides were synthesised on an Applied Biosystems 381A DNA synthesiser or were ordered through Pharmacia's custom oligonucleotide synthesis service.

Polyallomer Quick-seal ultracentrifuge tubes were supplied by Beckman Instruments Inc., USA and X-ray film by Fuji Photo Film Co., Japan.

All other reagents (SLR, AnalaR and Aristar grades) were supplied by BDH Ltd., Poole, UK and Fisons, Loughborough, UK.

DNA and amino acid sequence analysis was carried out using the GCG Sequence Analysis Software Package Version 6.0 (Devereux *et al.*, 1984).

2.0.1 Bacterial strains, culture conditions and plasmids

Haloferax volcanii (DSM 3757^T) was obtained as a freeze-dried culture from Deutsche Sammlung von Mikrooganismen und Zellkulturen, Germany. *Hf.volcanii* was grown in liquid batch culture in DSM medium 372 (200 g NaCl, 20 g $MgSO_4.7\ H_2O$, 2g KCl, 36 mg $FeCl_2.4\ H_2O$, 0.36 mg $MnCl_2.4H_2O$, 3 g trisodium citrate, 1 g sodium glutamate, 5 g yeast extract and 5 g casamino acids per litre, pH 7.2.) at 37°C, shaken at 180 rpm. Frozen stocks were maintained at -70°C in the same medium containing 15% glycerol.

Escherichia coli K12 HB101 F^- pro^- leu^- thi^- Sm^r $endoI^-$ $recA^-$ r_K^- m_K^- and *E.coli* K12 JM105 thi rpsL endA sbcB15 hspR4 Δ(lac-proAB)(F' traD36 proAB $lacI^q$ Z M15) was grown in 2xYT medium (16 g tryptone, 10 g yeast extract and 5 g NaCl per litre) at 37°C, shaken at 200 rpm. When required the medium was solidified by addition of 1.5% agar. Frozen stocks were maintained at -20°C and -70°C in the same medium containing 30% glycerol.

Ampicillin resistance plasmids were maintained or selected for in solid and liquid media by addition of ampicillin to a concentration of 100 μg ml^{-1}.

2.1 General methods

2.1.1 Manipulation of *Hf.volcanii*

2.1.1.1 Culture of *Hf.volcanii*

A freeze-dried culture was used to inoculate 50 ml of DSM medium 372 (Section 2.0.1) and grown for 5 days. Subsequent 1 litre cultures were inoculated with 5 ml of growing cells. After inoculation, 1 litre cultures were grown for 5-7 days. Cells were harvested by centrifugation at 8000g for 10 minutes in a Sorvall GSA rotor. Typical yield was 7-8 g litre^{-1} (cell pellet wet weight). The cell pellets were stored in sterile tubes at -20°C until required.

Plates were prepared by addition of 1.5% (w/v) agar to DSM medium 372. Colonies generally appeared 5 - 10 days after plating out cells. Plates were stored at 4°C for up to 4 weeks, or until the medium started to become dry and crystalline. When growing liquid cultures from plates a single colony was used to inoculate 20 ml of DSM medium 372.

2.1.1.2 Preparation of cell extract from *Hf.volcanii*

Cell pellets were resuspended (approximately 0.2 g ml^{-1}) in 20 mM Tris-(hydroxymethyl)-methylamine (Tris) -HCl (pH 8.0), 2 mM (disodium) ethylenediaminetetraacetate (EDTA) containing the required concentration of KCl or glycerol. The suspension was sonicated on ice in an MSE 150 Watt Ultrasonic Disintegrator Mk2 using a 9.5 mm probe. Four to six 20-second bursts were used, with 30 seconds cooling time between each. Cell debris

was removed by centrifugation at 10,000g for 20 minutes in a Sorvall SS-34 rotor.

2.2 Protein techniques

2.2.1 Chromatographic techniques

2.2.1.1 FPLC Superdex 200 gel filtration

The column was run in 20 mM Tris-HCl (pH 8.0), 2 mM EDTA, 2 M KCl at 1 ml min^{-1}. The enzyme was loaded using a 2 ml loop and 1 ml fractions were collected after the void volume (30 ml).

2.2.1.2 Matrex Gel Red A column

Columns of 10 ml bed volume were prepared in 10 ml disposable syringes plugged with siliconised glass wool. These columns were found to run freely under gravity at 1-2 ml min^{-1}. A column was first equilibrated with 20 mM Tris-HCl (pH 8.0), 2 mM EDTA, 0.1M KCl. Cell extract (10 ml) in the same buffer was passed through a 0.45 µm filter and loaded onto the column. The column was washed with 150 ml equilibration buffer to remove unbound proteins. Citrate synthase was eluted with 20 ml of 0.2 mM coenzyme A, 1 mM oxaloacetate in the same buffer, collecting in 2 ml fractions.

After each use the column was cleaned with 30 ml of 5 M urea, 0.5 M NaOH and washed with 200 ml Milli Q water before re-equilibrating.

2.2.5 SDS polyacrylamide gel electrophoresis

Pre-mixed Protogel 30% acrylamide, 0.8% N N' -methylene bisacrylamide was used. Acrylamide solution, gel buffer concentrate and water were mixed according to the table below and polymerised by the addition of ammonium persulphate and TEMED to a final concentration of 0.07% (w/v) and 0.075% (v/v) respectively.

Stacking gel buffer concentrate was 0.49 M Tris-HCl (pH 6.8), 0.4% (w/v) SDS. Resolving gel buffer concetrate was 1.50 M Tris-HCl (pH 8.9), 0.4% (w/v) SDS. These were mixed according to the table below:

		Resolving gel		Stack gel
	8%	10%	12.5%	
Acrylamide solution/ml	13	16	2	1.8
Gel buffer concentrate/ml	12	12	12	4.8
distilled water/ml	23	20	16	7.2

Samples were diluted two-fold in 0.125 M Tris-HCl (pH 6.8), 0.04% (w/v) SDS, 20% (w/v) sucrose, 10% (v/v) β-mercaptoethanol, 0.04% (w/v) bromophenol blue and boiled for 2 minutes prior to loading on the gel.

Gels were run in 0.052 M Tris, 0.001% (w/v) SDS, 0.4% (w/v) glycine. Small (60 x 100 x 1 mm) gels were run on an Atto Corp. Mini-Atto system at 10 mA (stacking gel) and 20 mA (resolving gel). Large (160 x 160 x 2 mm)

gels were run on a Bio-Rad Protean II xi system at 20 mA (stacking gel) and 30 mA (resolving gel). All were run at constant current.

The protein standards used were from Biorad (low molecular weight range): Phosphorylase b (M_r 97,400), Bovine serum albumin (M_r 66,200), Ovalbumin (M_r 45,000), Carbonic anhydrase (M_r 31,000), Soybean trypsin inhibitor (M_r 21,500) and Lysozyme (M_r 14,400).

Gels were stained in 45% (v/v) methanol, 10% (v/v) acetic acid, 0.25% (w/v) Coomassie Blue R for 30-60 minutes and destained in 5% (v/v) methanol, 7.5% (v/v) acetic acid.

2.2.6 Western blotting

A large polyacrylamide gel was run as described above, although the gel was not stained. A 100 x 100 mm region containing the protein bands was cut from the gel and a blotting sandwich prepared with this region alone. The blotting sandwich was set up in the following order:

Five 100 x 100 mm sheets of Whatman 3MM paper soaked in blotting buffer (50 mM Tris, 50 mM glycine, 20% (v/v) methanol), the gel section, two 100 x 100 mm sheets of Immobilon P (Polyvinylidene difluoride) membrane (pre-wetted in 100% methanol and then soaked in blotting buffer for 10 minutes) and a further five 100 x 100 mm sheets of Whatman 3MM paper also soaked in blotting buffer.

The sandwich was placed in a blotting tank containing 3.5 litres of blotting buffer pre-cooled to 4°C and electrophoretically blotted at 4°C for 15 hours at 200 mA constant current.

The gel was then removed and stained with Coomassie Blue R as described above. The membranes were stained for 15 minutes in 50%

methanol, 10% (v/v) acetic acid, 0.025% (w/v) Coomassie Blue R and destained for 20-30 minutes in several changes of 50% (v/v) methanol, 10% (v/v) acetic acid. The membranes were then washed 3 times in Milli Q water for 20 minutes, sealed in bags while still wet and stored at -20°C.

2.2.7 Citrate synthase assays

Citrate synthase activity was measured spectrophotomerically by the method of Srere *et al.* (1963). Free coenzyme A reacts with 5,5'-dithio bis-2-nitrobenzoate (DTNB), releasing thionitrobenzoate which has an absorbtion maximum at 412 nm and a molar extinction coefficient of 13,600 l $mol^{-1}cm^{-1}$.

Assays were carried out at 30°C in a Perkin Elmer Lambda 3B spectrophotometer. Initial rates were calculated using the PECSS software supplied with the instrument. 1 unit of enzyme activity is defined as the amount which produces 1 µmol of product min^{-1}.

Unless otherwise specified, assays were carried out in 1 ml of 20 mM Tris-HCl (pH 8.0), 2 mM EDTA, 2 M KCl containing 0.2 mM oxaloacetate, 0.4 mM acetyl coenzyme A and 0.1 mM DTNB.

2.2.8 Acetylation of coenzyme A

Coenzyme A was acetylated by dissolving 10 mg CoA in 1 ml of Milli Q water, cooling on ice and adding 0.2 ml of 1 M $KHCO_3$, followed by 0.1 ml of 1 M acetic anhydride freshly diluted with Milli Q water. This gives a seven-fold molar excess of acetic anhydride over CoA. After incubation on ice for 10 minutes, acetylation was checked by addition of a 20 µl sample to assay buffer containing DTNB.

2.3 DNA methods

2.3.1 Quantitation of DNA

DNA concentrations were determined spectrophotometrically by measurement of absorbance at 260 and 280 nm. An absorbance of 1 at 260 nm corresponds to approximately 37 µg (single-stranded) or 50 µg (double-stranded) DNA. An A_{260}/A_{280} ratio of greater than 1.8 was taken to indicate purity (Maniatis *et al.*, 1982).

2.3.2 DNA preparation methods

2.3.2.1 Preparation of genomic DNA from *Hf.volcanii*

Genomic DNA was prepared by a modification of the method of Holmes & Dyall-Smith (1991).

15 ml of culture was transferred to a sterile 50 ml tube and spun at 3,000 rpm for 10 minutes in a benchtop centrifuge to pellet the cells.

The pellet was resuspended in 6 ml of TE pH 8.0 containing 0.5% SDS, 100 µg ml^{-1} proteinase K and incubated at 37°C for 1 hour.

1 ml of 5.0 M NaCl was added and mixed thoroughly, followed by 0.8 ml of 8% (w/v) hexadecyltrimethyl ammonium bromide (CTAB), 0.6 M NaCl. The mixture was incubated at 65°C for 10 minutes.

The supernatant was removed to a clean, sterile tube and extracted with an equal volume of chloroform/isoamyl alcohol for 10 minutes.

The phases were separated by spinning at 2000 rpm for 5 minutes and the supernatant transferred to a clean, sterile tube. 0.6 Volumes of

isopropanol were added to the supernatant. The precipitated DNA was transferred to 70% (v/v) ethanol by spooling onto a sterile glass rod.

The DNA was washed with 70% (v/v) ethanol, allowed to air-dry and dissolved in 2 ml of TE pH 8.0.

2.3.2.2 Plasmid DNA 'Maxiprep' method I

Log phase *E.coli* culture (1 ml) was used to inoculate 500 ml 2xYT, containing ampicillin (100 μg ml^{-1}) which was then grown overnight. The cells were spun down for 10 minutes at 8,000g in a Sorvall GSA rotor. The pellet was resuspended in 2 ml of 25 mM Tris-HCl (pH 8.0), 50 mM glucose, 10 mM EDTA and transferred to a 30 ml Corex tube.

NaOH (4 ml of 0.2 M) containing 1% (w/v) SDS was added, mixed by inversion and incubated on ice for 5 minutes. Ice-cold potassium acetate (3 ml of 3 M, pH 4.8) was added, mixed by inversion and incubated on ice for 10 minutes.

Cell debris was removed by spinning in a Sorvall SS-34 rotor at 8,000g for 20 minutes at 4°C. The supernatant was made up to 10 ml with distilled water and the DNA purified by caesium chloride density gradient cetrifugation.

2.3.2.3 Caesium chloride equilibrium density gradient centrifugation

To the 10 ml of DNA solution prepared by the 'Maxiprep' method I, 10.17 g caesium chloride was added. This was sealed in a 13 x 51 mm Beckman Quick-Seal ultracentrifuge tube and spun at 55,000 rpm for 24 hours at 4°C in a Beckman Ti70 rotor.

The plasmid DNA was recovered under UV illumination using a syringe and 19G needle and extracted three times with an equal volume of water-saturated butanol to remove the ethidium bromide. The DNA was then precipitated with ethanol, washed in 70% (v/v) ethanol, dried and dissolved in 0.5 ml of TE (pH 8.0).

2.3.2.4 Plasmid DNA 'Maxiprep' method II

Cells were grown overnight in 100 ml 2xYT, containing ampicillin (100 $\mu g\ ml^{-1}$) after inoculation with 1 ml of log phase *E.coli* culture.

The cells were spun in a benchtop centrifuge at 3,000 rpm for 10 minutes and the pellet resuspended in 4 ml of 25 mM Tris-HCl (pH 8.0), 50 mM glucose, 10 mM EDTA. NaOH (8 ml of 0.2 M) containing 1% (w/v) SDS was added, mixed by inversion and incubated on ice for 10 minutes. Ice-cold potassium acetate (6 ml of 3 M, pH 4.8) was added, mixed by inversion and incubated on ice for 15-30 minutes. Cell debris was removed by spinning at 3,000 rpm for 15 minutes in a benchtop cetrifuge.

The supernatant was added to 18 ml of propan-2-ol and incubated on ice for 30 minutes. The precipitated nucleic acid was pelleted in a benchtop centrifuge at 3,000 rpm for 20 minutes. The pellet was dissolved in 1 ml of TE (pH 8.0), 1 ml of 6 M LiCl was added and incubated on ice for 15 minutes. The precipitate was removed by spinning in a microcentrifuge for 10 minutes. The supernatant was added to 4 ml of absolute ethanol and the nucleic acids precipitated at -20°C for 30 minutes.

The precipitate was pelleted in a microcentrifuge for 10 minutes, washed with 70% (v/v) ethanol, dried and dissolved in 0.4 ml of TE (pH 8.0). To the solution, 4 µl of 10 $mg\ ml^{-1}$ RNAse A was added and the mixture

incubated at 37°C for 30 minutes. To inactivate the RNAse 20 µl of 10% (w/v) SDS was added, the solution was heated to 75°C for 10 minutes, 0.42 ml of 6 M LiCl was added and mixture incubated at room temperature for 15 minutes. The precipitate was pelleted in a microcentrifuge for 10 minutes and the supernatant added to 2 ml ethanol containing 40 µl of 3M sodium acetate (pH 5.2); DNA was precipitated at -20°C for 30 minutes.

The DNA was pelleted in a microcentrifuge for 10 minutes, washed with 70% (v/v) ethanol, dried and dissolved in 0.4 ml of distilled water. The DNA solution was extracted once with phenol / chloroform and twice with chloroform and ethanol precipitated. The precipitate was pelleted in a microcentrifuge for 10 minutes, washed with 70% (v/v) ethanol, dried and dissolved in 0.1 ml of distilled water.

This procedure yields 400-800 µg plasmid DNA suitable for restriction digests or sequencing, without CsCl density gradient centrifugation.

2.3.2.5 Plasmid DNA 'Minipreps'

Small-scale isolation of plasmid DNA was carried out by the alkaline lysis method of Maniatis *et al.*(1982).

2.3.2.6 Plasmid DNA 'Magic Minipreps'

Plasmid DNA was isolated using the Magic Minipreps kit, following the manufacturer's instructions.

2.3.3 Restriction digests of DNA

DNA was digested with the appropriate enzyme under the buffer conditions specified by the manufacturer. When the DNA was digested with

two enzymes requiring different buffer concentrations, the first enzyme was heat inactivated under the conditions recommended by the supplier and further 10x buffer was added to the reaction mixture prior to addition of the second enzyme.

2.3.4 Agarose gel electrophoresis of DNA

Agarose gels containing 0.6-1.0% (w/v) agarose were made with 1x Tris-acetate-EDTA (TAE) buffer containing 0.5 μg ml^{-1} ethidium bromide. Loading buffer (6x concentration) was 40% (w/v) sucrose, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol. Gels were run at 6 Vcm^{-1} (Plasmid DNA and plasmid digests) or 2 Vcm^{-1} (genomic DNA digests) in 1x TAE buffer. When the DNA bands were to be cut from the gel and purified, ultra-pure low-melting point agarose was used.

2.3.5 Recovery of DNA from agarose gels

DNA was recovered from low-melting point agarose gels by a modification of the freeze-sqeeze method of Tautz & Renz (1983). The gel slice was placed in a 0.5 ml Eppendorf tube which had been pierced at the base and the hole stopped with siliconised glass wool. The tube was frozen in liquid nitrogen for 30 seconds, placed into a 1.5 ml Eppendorf tube and spun in an micro centrifuge for 5 minutes. The eluent containing the DNA was collected.

2.3.6 Purification of DNA fragments

Phenol was equilibrated with Tris-HCl (pH 8.0), phenol / chloroform / isoamyl alcohol and chloroform / isoamyl alcohol extractions were performed as described by Maniatis *et al.*(1989).

DNA was purified using NENsorb cartridges according to the manufacturer's instructions, eluting with 30% (v/v) ethanol. Aristar grade methanol was used to wet the columns. Alternatively, the Geneclean II kit was used to purify DNA fragments, according to the manufacturer's instructions.

Spun columns of Sepharose CL-6B, Sephadex G-25 and Sephadex G-50 were prepared in 0.5 ml Eppendorf tubes which had been pierced at the base and the hole stopped with siliconised glass wool. 400 µl of 50% slurry was packed by spinning inside a 1.5 ml Eppendorf tube. Four column volumes of TE (pH 8.0) were passed through by spinning under the same conditions and the DNA sample was loaded onto the dry column and re-spun into a fresh tube.

2.3.7 Precipitation of DNA

DNA precipitations were carried out using ethanol and sodium acetate as described by Maniatis *et al.*(1982).

2.3.8 Ligation procedures

Ligation of cohesive ends was carried out using 0.1 Units of Gibco BRL T4 DNA ligase in the buffer supplied by the manufacturer. The ratio of insert to vector was approximately 5 to 1, estimated by ethidium bromide

staining on agarose gel. Ligation was carried out overnight at 15°C in a total volume of 20 µl.

The Gibco T4 ligase buffer is suitable for blunt-end ligations, which were therefore carried out under the same conditions except that the ratio of insert to vector was increased to approximately 10 or 20 : 1.

2.3.9 DNA blotting techniques

2.3.9.1 Southern blotting

Restriction digests of plasmid DNA were carried out as described. Digests were run on 1% (w/v) agarose gels in 1x TAE buffer at 2 - 6 Vcm^{-1}, dependent upon the size of the DNA fragments. The size markers used were 2 µg of λ *Hin*dIII digested DNA.

Blotting onto Genescreen membrane was carried out by the capilliary method (Southern, 1975). The DNA was first denatured by incubation in 0.2 M NaOH, 0.6 M NaCl for 30 minutes at room temperature. The gel was then washed three times in 10x SSC (1.5 M NaCl, 0.15 M sodium citrate) for 20 minutes at room temperature. The Genescreen membrane was soaked in 10x SSC for 10 minutes and the blotting stack was set up using a reservoir containing 10x SSC. Blotting was carried out for 18 hours. The membrane was removed and air dried. The DNA was baked onto the membrane at 80°C for 3 hours.

2.3.9.2 Dot blotting

The DNA sample was dissolved in water and spotted onto dry Genescreen membrane using a fine-tipped pipette. Generally 1 µg of DNA was applied in a single dot. The dot was allowed to dry and then the blot was

treated with 1.5 M NaCl, 0.5 M NaOH and 1.5 M NaCl, 0.5 M Tris-HCl (pH 7.2), 1 mM EDTA as described in 2.3.9.3 below. The blot was placed on a filter-paper pad soaked in 2x SSC for 2 minutes, air-dried and baked at 80°C for 2 hours.

2.3.9.3 Colony hybridisation

Gridded Hybond-N nylon membranes were placed on the surface of 2xYT plates containing ampicillin (100 μg ml^{-1}). Bacteria were streaked onto the membranes and grown overnight at 37°C. The membranes were placed on a pad of filter paper soaked in 1.5 M NaCl, 0.5 M NaOH for 7 minutes. They were then transferred to a fresh pad soaked in 1.5 M NaCl, 0.5 M Tris-HCl (pH 7.2), 1 mM EDTA for 3 minutes. This last step was repeated once and the membranes then washed in 2x SSC before being allowed to air-dry. The DNA was baked onto the membranes at 80°C for 2 hours.

2.3.10 Deprotection and purification of synthetic oligonucleotides

Columns from the DNA synthesiser were opened and the resin removed to a 2 ml screw-cap tube containing 1ml concentrated ammonia. The resin was resupended and incubated at 55°C for 18 hours to cleave the DNA from the column and remove protecting groups.

The solution was cooled on ice, centrifuged to pellet the resin and the supernatant containing the DNA removed. The solution was dried under vacuum, dissolved in 0.5 ml distilled water, centrifuged to remove any remaining resin and precipitated with ethanol. The pellet was redissolved as

before and precipitated a second time before finally dissolving in 0.5 ml of sterile Milli Q water.

2.3.11 End-labelling of oligonucleotides

Oligonucleotides were end-labelled with [γ-^{32}P] ATP by incubating 20 pmoles oligonucleotide with 100 µCi [γ-^{32}P] ATP (33 pmoles) and 10 units of T4 polynucleotide kinase in 1x kinase buffer (20 mM Tris-HCl (pH 7.6), 5 mM $MgCl_2$, 5 mM dithiothreitol, 50 µg ml^{-1} bovine serum albumin (BSA)) at 37°C for 30 minutes. Total reaction volume was 40 µl.

Incorporation of radiolabel was checked and unincoporated label removed by passing the solution through a Sephadex G-25 spun column (for 17-20mers) or a G-50 spun column (for larger olignucleotides), as described under DNA purification methods.

2.3.12 Hybridisation of labelled oligonucleotide probes to DNA blots

The dried membrane was sealed in a bag containing 50 ml pre-hybridisation solution (5.2x SSC, 5x Denhardt's solution [0.1% (w/v) Ficoll, 01% (w/v) polyvinylpyrrolidone, 0.1% (w/v) BSA], 0.1% (w/v) , 0.6% (w/v) SDS, 20 µg ml^{-1} salmon sperm DNA). The salmon sperm DNA was boiled for 5 minutes and then cooled on ice for 5 minutes prior to addition. The membrane was incubated overnight at the desired temperature. To this, 20 pmoles of labelled oligonucleotide probe were added and the membrane incubated for a further 12-18 hours.

The membrane was washed twice with 50 ml of 2x SSC, 0.1% (w/v) SDS at the same temperature, sealed in a plastic bag and autoradiographed

using pre-flashed X-ray film and intensifying screens at -70°C. Two pieces of film were used, the first developed after 2 days, the second after 4-30 days depending on the intensity of the bands observed on the first.

2.3.13 Preparation of competent *E.coli* JM103 and HB101

Competent *E.coli* were prepared by the calcium chloride method. Growth medium (50 ml of 2xYT in a 250 ml flask) was inoculated with 1 ml of an overnight bacterial culture and grown at 37°C for 2 hours with vigorous shaking. The culture was chilled on ice for 10 minutes and spun at 2,500 rpm in a benchtop centrifuge for 5 minutes.

The supernatant was discarded and the cell pellet resuspended in 50 ml of ice-cold, sterile 50 mM $CaCl_2$. The cell suspension was kept on ice for 30 minutes and the centrifugation repeated. The cells were resuspended in 5 ml of ice-cold 50 mM $CaCl_2$ and kept on ice for 60 minutes before use.

2.3.14 Transformation of competent *E.coli* JM103 and HB101 with plasmid DNA

Competent cells (200 µl) were added to 5 µl of a ligation mixture and incubated on ice for 30 minutes in polypropylene tubes. The cells were heat-shocked at 42°C for 90 seconds and replaced on ice for 60 seconds. The culture was made up to 1 ml by addition of 800 µl of 2xYT and the cells incubated at 37°C for 1 hour without shaking. X-Gal (50 µl of 20 mg ml^{-1}, in DMF) and IPTG (50 µl of 24 mg^{-1} ml^{-1}) were added and 200-350 µl of the mixture plated on 2xYT agar containing ampicillin (100 µg ml^{-1}). Plates were incubated at 37°C overnight.

2.3.15 PCR techniques

2.3.15.1 PCR amplification of genomic DNA

Amplification was carried out using NBL and Gibco BRL Taq polymerases in 100 µl of 20mM Tris-HCl (pH 8.3), 25 mM KCl, 0.5-5 mM $MgCl_2$, 0.1 mg ml^{-1} BSA overlaid with 50 µl of mineral oil. Reactions were carried out in a Perkin Elmer Cetus DNA Thermal Cycler using Perkin Elmer GeneAmp reaction tubes.

Reaction mixtures contained 80 ng of target DNA, 50 µM dATP, deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP) and deoxythymidine triphosphate (dTTP), primers at a concentration of 0.4 µM each and 2.5 U of Taq polymerase. The mixture was heated to 94°C for 4 minutes, maintained at 94°C while the polymerase was added and then 30 reaction cycles were carried out using the following profile:

- 94°C for 1 minute
- 37°C-65°C for 1.5 minutes
- 72°C for 2 minutes.

A final 10 minute extension was run after the last cycle was complete. A 10 µl aliquot from each reaction was run on a 1% (w/v) agarose gel.

2.3.15.2 Asymmetric PCR

The technique of asymmetric PCR described by McCabe (1990) was used to generate single-stranded DNA by reamplification of a gel-purified PCR product.

Reactions were carried out in 100 µl of 20 mM Tris-HCl (pH 8.3), 25 mM KCl, 1.5 mM $MgCl_2$, 0.1 mg ml^{-1} BSA, overlaid with 50 µl of mineral oil. The ratio of primers was 50 pmoles to 1pmole. The primer to be extended into the single-stranded product was present in the greater amount.

The template was prepared by running 10 µl of a previous PCR reaction on a 1% (w/v) agarose gel, excising the band to be reamplified, recovering the DNA by freeze-squeeze followed by ethanol precipitation or using Geneclean and finally dissolving the dried pellet in 20 µl distilled water. Of this, 0.5 µl was used for each reamplification reaction.

The reactions were carried out as described above, the mineral oil was removed by extraction with chloroform and an equal volume of 4 M ammonium acetate was added to the aqueous phase. DNA was precipitated by adding isopropanol to a final concentration of 50% (v/v) and leaving the solution for 10 minutes at room temperature. The precipitate was pelleted by spinning for 10 minutes in a microcentrifuge, washed with ice cold 70% (v/v) ethanol and dried.

The pellet was dissolved in 7 µl of distilled water and sequenced using 1 pmole of the PCR primer that was at the lower concentration in the amplification reaction. The sequencing procedure was identical to that for plasmid DNA, except that the denaturing step was omitted.

2.3.16 Sequencing plasmid DNA

Sequencing of double-stranded template DNA (pUC18) was carried out by the dideoxy chain termination method, using Sequenase kits and α-^{35}S-dATP (1000Ci/mmol), according to the manufacturer's instructions. For each set of 4 reactions 3-5 µg of template DNA were used. Between 1 and

20 pmoles of primer were used per set of reactions. When sequencing close to the primer 1 µl of Sequenase Mn^{2+} buffer was added to the reaction mixture.

Sequencing reactions were run in Tris-borate-EDTA (TBE) buffer on 0.4 mm thick 8% polyacrylamide gels made using Sequagel reagent kits. Gels were left 2-3 hours after pouring before use and were pre-run for 1 hour. Using a Bio-Rad Sequi-Gen II 50 x 21cm sequencing cell, gels were run at 40-50 Watts constant power and at a constant temperature of 50°C. Gels were fixed in 10% (v/v) acetic acid, 10% (v/v) methanol for 15 minutes before drying under vacuum.

Dried gels were autoradiographed overnight, or longer if necessary.

Chapter 3

Characterisation of activity and stability of *Hf.volcanii* citrate synthase

3.1 Introduction

The effect of monovalent salt concentration on the activity and stability of *Hf.volcanii* citrate synthase was studied under different conditions as a basis for the design of improved purification procedures. The effect of KCl concentration on stability was investigated by thermal inactivation experiments. These serve as the first steps in the thorough characterisation of the enzyme that will be necessary if it is to be used as a model for the study of halophilicity.

The studies described below were carried out on cell extracts prepared as described in Section 2.1.1.2. When citrate synthase assays were carried out on unfractionated extracts, background activities of coenzyme A hydrolysis were measured by assays omitting oxaloacetate. In cases where this rate was not zero, the background value was subtracted from the measured rate to give the citrate synthase activity.

The apparent K_m values of the *Hf.volcanii* citrate synthase for oxaloacetate and coenzyme A have been determined as 38 µM and 75 µM respectively in 2M NaCl (Danson *et al.*, 1985). This was taken as a guide to the enzyme's kinetic characteristics in 2M KCl: assays were therefore carried out in 200 µM oxaloacetate and 400 µM coenzyme A.

3.2 Salt dependence of citrate synthase activity

Cell extracts were prepared in 20 mM Tris-HCl (pH 8.0), 2 mM EDTA, containing either 0.5 M NaCl or KCl. Samples of 10 µl were assayed at increasing final concentrations of NaCl or KCl as shown in Figure 3.1. Each assay was carried out in triplicate.

3.3 Stability of citrate synthase

Cell extracts were prepared in 20 mM Tris-HCl (pH 8.0), 2 mM EDTA, containing either 0.5 M NaCl or KCl. Samples of 90 µl were diluted into 3 ml 20 mM Tris-HCl (pH 8.0), 2 mM EDTA, containing NaCl or KCl, to give final concentrations of 0, 0.05, 0.1, 0.5, 1, 2, 3 and 4 M salt. The enzyme solution was incubated at 25°C and 30 µl samples were assayed for citrate synthase activity in buffer containing 2 M NaCl or KCl after 0, 24, 48 and 120 hours.

It was found that at NaCl or KCl concentrations below 0.1 M, total inactivation occurred within 24 hours. At higher concentrations of either salt no activity was lost during the 120 hour course of the incubation (data not shown).

3.4 Thermal inactivation of citrate synthase

Cell-free extracts were prepared in 20 mM Tris-HCl (pH 8.0), 2 mM EDTA containing 2 M KCl. A glass tube containing 2.85 ml of the same buffer was heated in a water bath and 150 µl of cell extract added. The solutions were mixed briefly with a 1 ml automatic pipette. Inactivations were carried out at 60, 62, 64, and 66°C. At higher temperatures the mixing

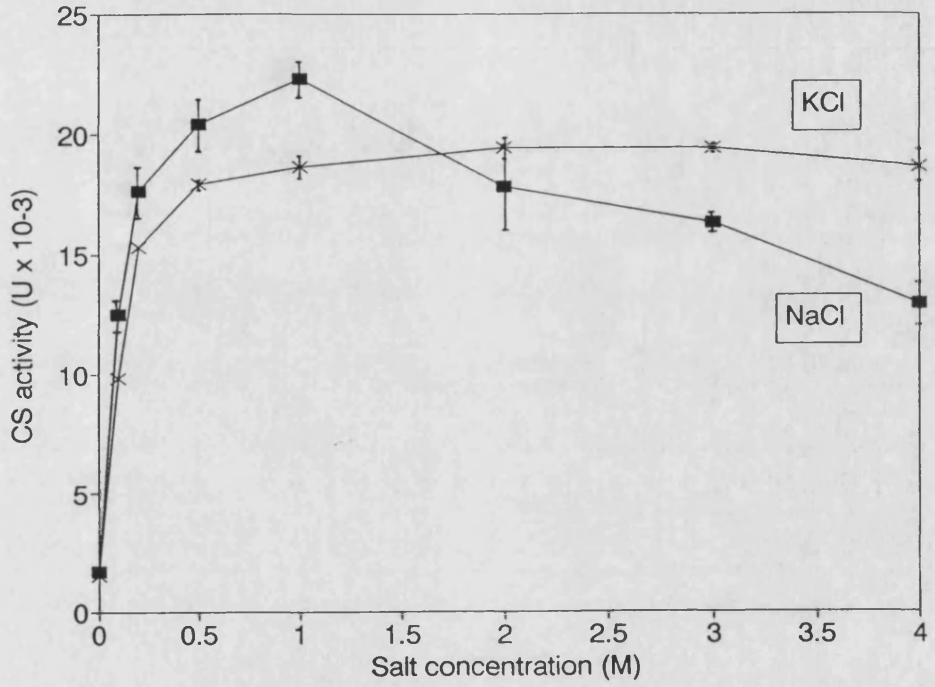

Figure 3.1 Dependence of citrate synthase activity on salt concentration

Initial rates are plotted against the concentration of NaCl or KCl in the assay. Each point is the mean value of triplicate assays: the error bars indicate 1 standard deviation.

process was found to produce transient cooling of the system, resulting in non-linear rates of inactivation. The effect became more marked with increasing inactivation temperatures.

Temperatures were measured using a thermocouple in a separate, identical tube and were maintained to within 0.2°C of the target temperature throughout the timecourse.

Aliquots of approximately 150 µl were removed at intervals, cooled rapidly by pipetting into ice-cold glass tubes and 50 µl of this sample were then assayed immediately. As it was not possible to measure hot liquids accurately with the air-displacement pipettes used, the rapid cooling of the aliquots by pipetting into an ice-cold glass tube was essential before removing the 50 µl for assay. A value for 100% activity was obtained by diluting 150 µl of cell-free extract into 2.85 ml of the same buffer at 4°C and assaying 50 µl aliquots. The rates of inactivation were determined by plotting the natural logarithm of the percentage initial activity as a function of time. The apparent first order rate constants are given by the slopes of these plots, which are shown in Figures 3.2-3.5.

According to transition-state theory, the relationship between the rate constant of a reaction and the absolute temperature can be expressed as:

$$k = k_B T/h \, .e^{-\Delta G^{\circ *}/RT}$$

Where k is the rate constant of the reaction, k_B is Boltzmann's constant, h is Planck's constant , R is the gas constant, T is the absolute temperature in degrees Kelvin and $\Delta G^{\circ *}$ is the free energy change in going from the ground state of the reactant(s) to the transition state (Creighton, 1993).

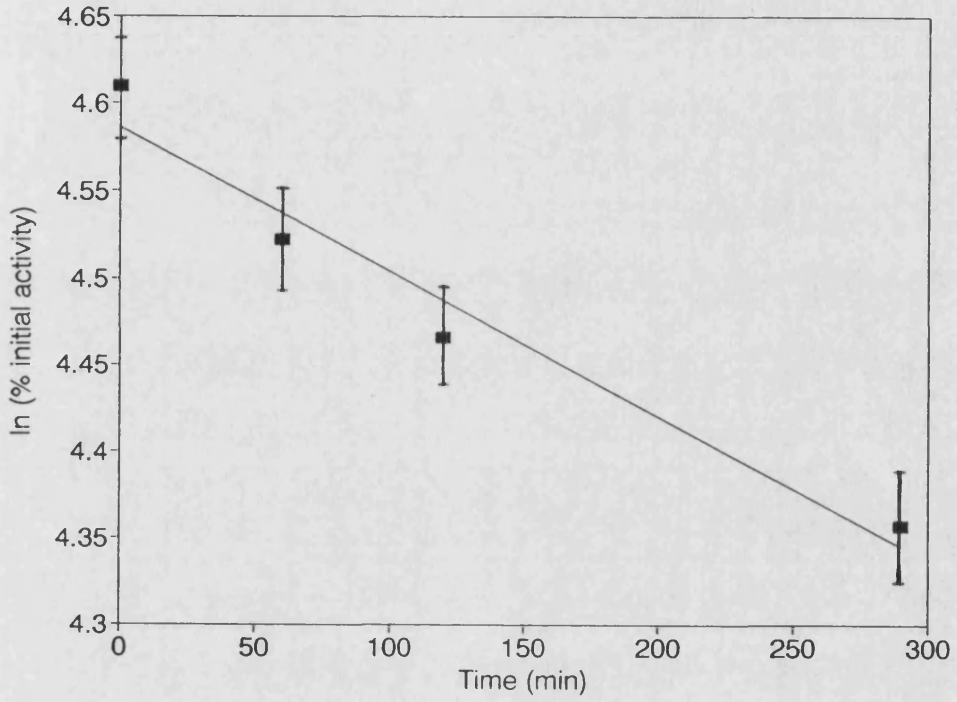

Figure 3.2 Thermal inactivation of citrate synthase at 60°C

The graph shows ln (% initial activity) plotted against time in minutes. The rate constant of the inactivation *k*, equal to -(slope), is 8.3×10^{-4} min^{-1}. Each point is the mean of triplicate independent inactivations; error bars indicate 1 standard deviation.

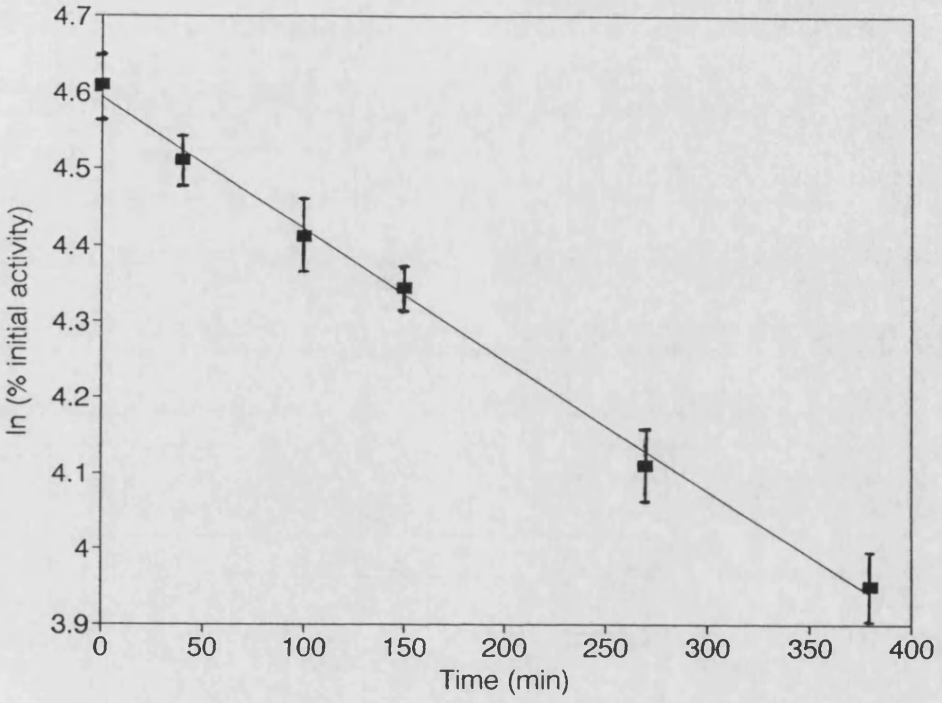

Figure 3.3 Thermal inactivation of citrate synthase at 62°C

The graph shows ln (% initial activity) plotted against time in minutes. The rate constant of the inactivation *k*, equal to -(slope), is 1.7×10^{-3} min^{-1}. Each point is the mean of triplicate independent inactivations; error bars indicate 1 standard deviation.

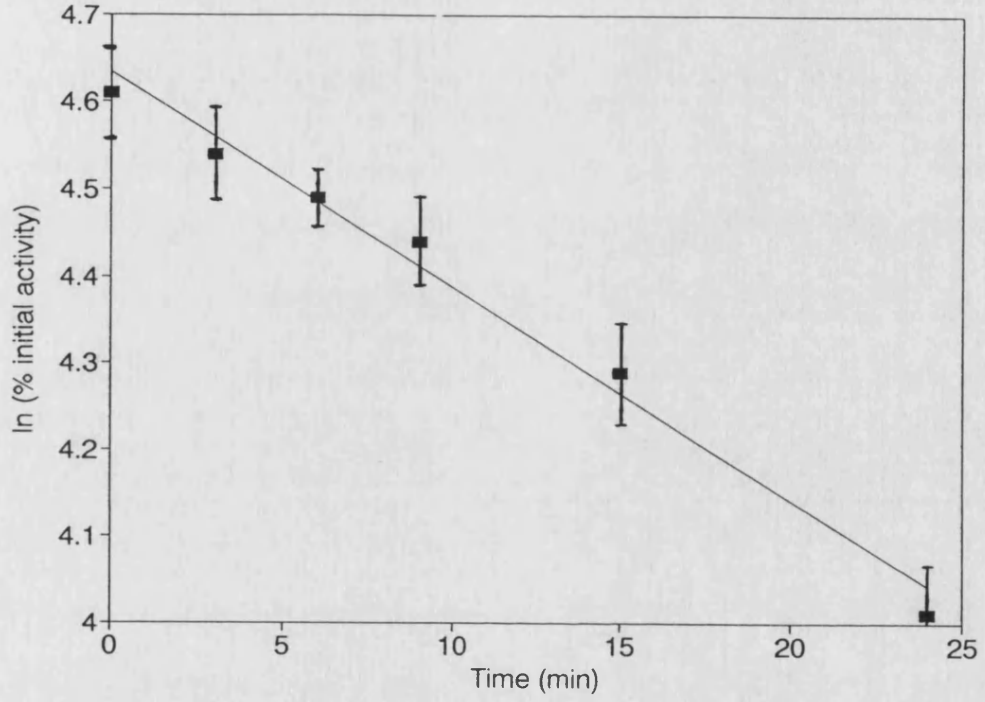

Figure 3.4 Thermal inactivation of citrate synthase at 64°C

The graph shows ln (% initial activity) plotted against time in minutes. The rate constant of the inactivation *k*, equal to -(slope), is 0.025 min^{-1}. Each point is the mean of triplicate independent inactivations; error bars indicate 1 standard deviation.

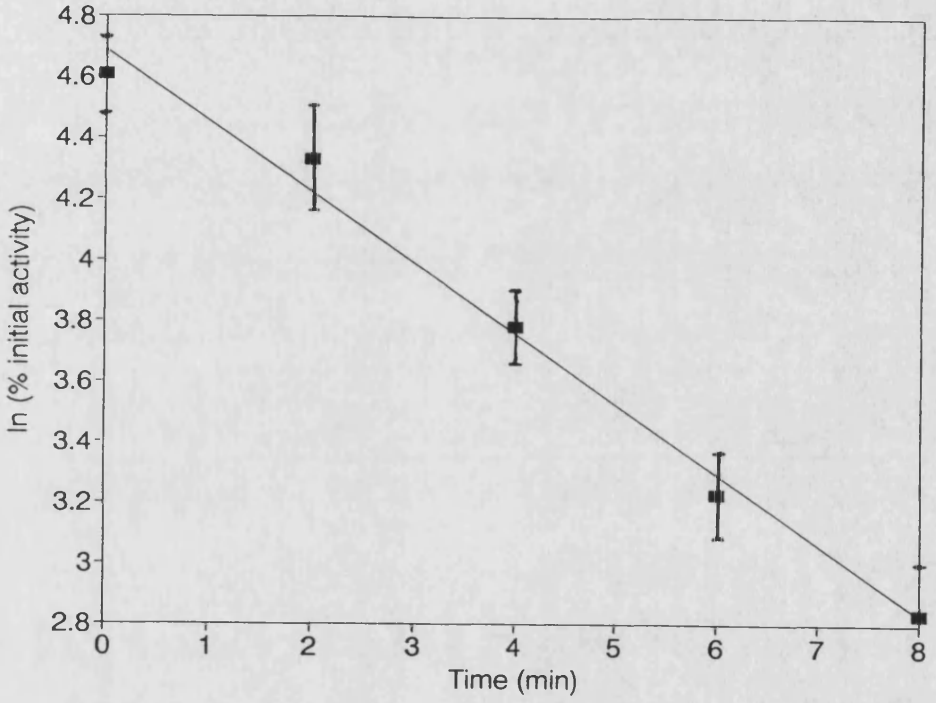

Figure 3.5 Thermal inactivation of citrate synthase at 66°C

The graph shows ln (% initial activity) plotted against time in minutes. The rate constant of the inactivation *k*, equal to -(slope), is 0.233 min^{-1}. Each point is the mean of triplicate independent inactivations; error bars indicate 1 standard deviation.

Another form of this equation may be obtained by substituting:

$$\Delta G^{\circ *} = \Delta H^{\circ *} - T\Delta S^{\circ *}$$

Hence:

$$k = k_B T/h \, .e^{-\Delta H^{\circ *}/RT} \, .e^{\Delta S^{\circ *}/R}$$

Over small temperature ranges $\Delta S^{\circ *}/R$ can be taken as a constant. Since, with any change in temperature, k varies proportionally to $k_B T/h$, whereas it varies exponentially with $-\Delta H^{\circ *}/RT$, the term $k_B T/h$ may be regarded as constant. Therefore the equation can be expressed:

$$\ln k = [\ln(k_B T/h) + \Delta S^{\circ *}/R] - \Delta H^{\circ *}/R \, . \, (1/T)$$

This shows that under these assumptions the natural logarithm of the rate constant is a linear function of the reciprocal of the absolute temperature.

The natural logarithm of each rate constant was plotted against the reciprocal of the appropriate inactivation temperature (in degrees Kelvin) to produce a straight line (An Arrhenius plot). This plot is shown in Figure 3.6a, while Figure 3.6b shows for comparison similar data obtained in our laboratory for the citrate synthases of *Sulfolobus solfataricus*, *Thermoplasma acidophilum* and pig heart (unpublished data).

Inactivations were carried out at 64°C in 20 mM Tris-HCl (pH 8.0), 2 mM EDTA containing 1.0, 1.6, 1.7, 1.8, 2.0, 2.2 and 3.0 M KCl, to investigate the effect of KCl concentration on the thermostability of the enzyme. The rate constants of inactivation were determined as above and the combined data are plotted in Figure 3.7. Inactivation in 1.0 M KCl was too rapid to measure, while no inactivation was observed in the presence of 3 M KCl.

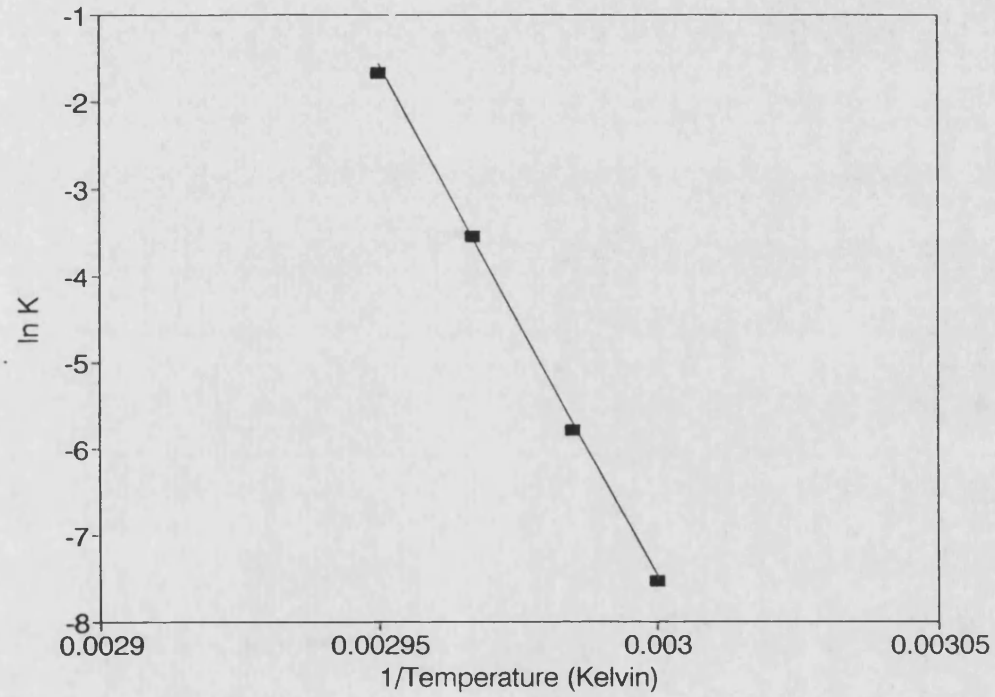

Figure 3.6a Arrhenius plot of thermal inactivation of citrate synthase

The graph shows ln (rate constant of thermal inactivation *k*) plotted against the reciprocal of the absolute temperature in Kelvin^{-1}. The slope of the graph, equal to $-\Delta H^*/R$, is - 117870 JK (standard error = 1960).

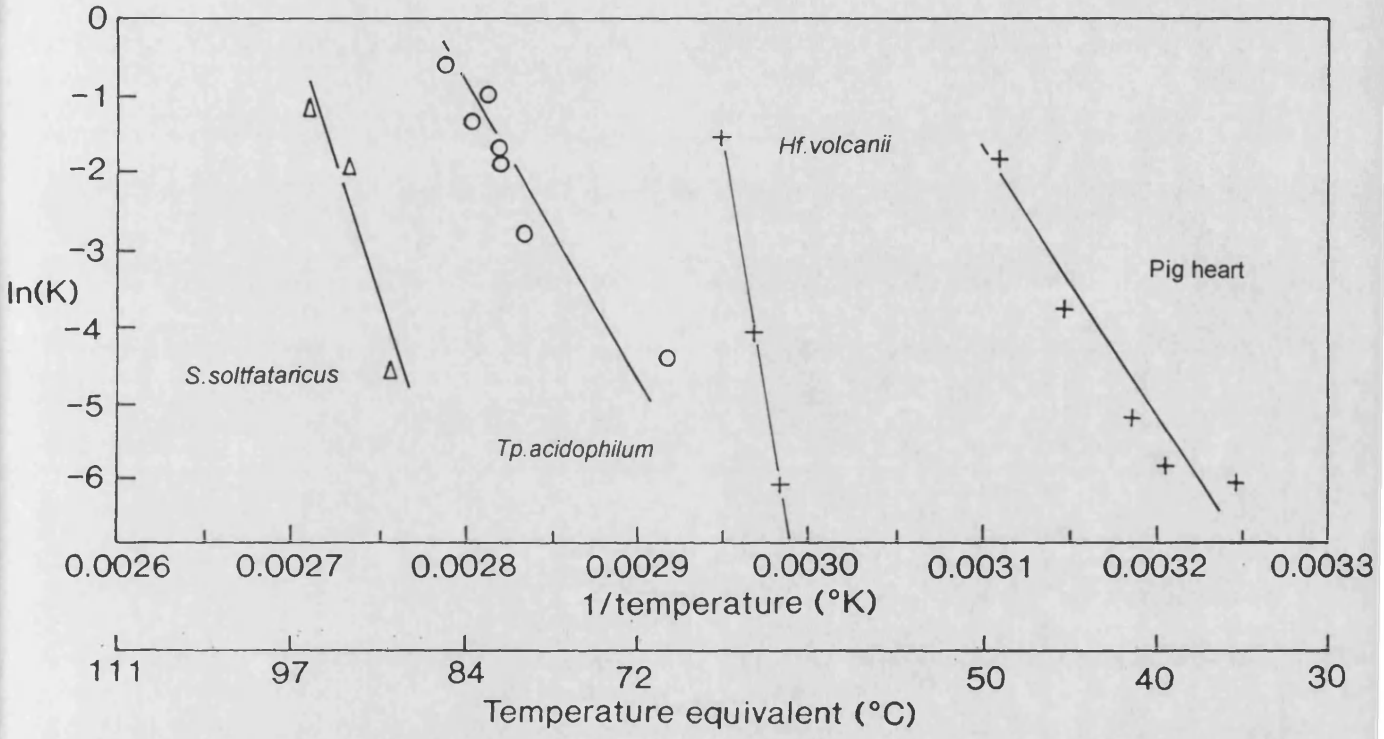

Figure 3.6b Arrhenius plot of thermal inactivation of citrate synthases

This graph compares the Arrhenius data for the *Hf.volcanii* citrate synthase with those obtained in our laboratory for citrate synthases from other sources (unpublished data).

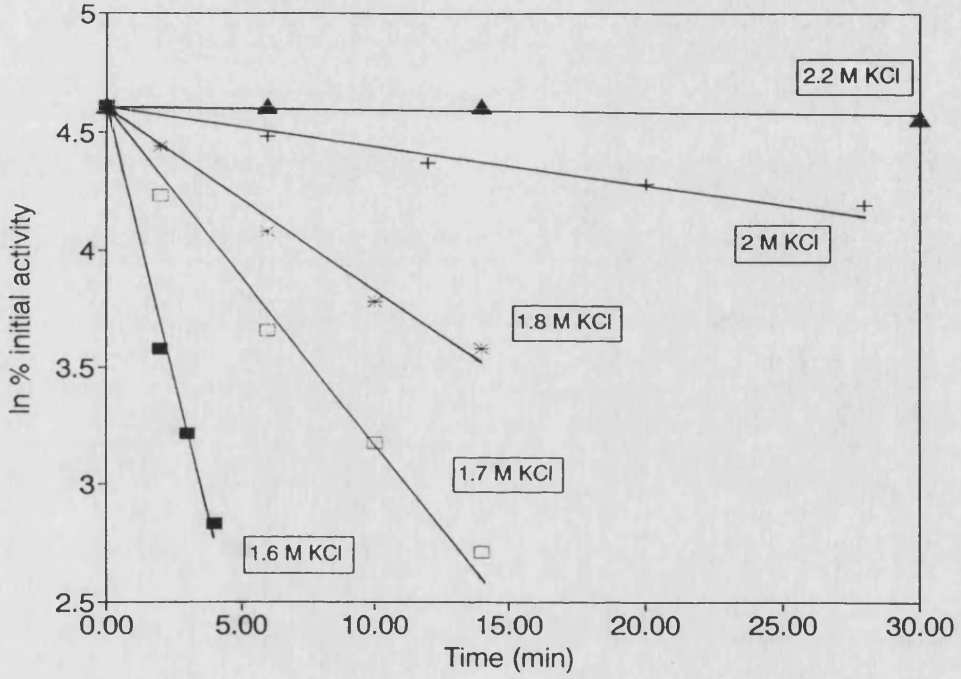

Figure 3.7 Effect of KCl concentration on thermal inactivation of citrate synthase at 64°C

The graph shows ln (% initial activity) plotted against time in minutes for a range of KCl concentrations.

3.5 Discussion

At 30-40°C *Hf.volcanii* grows optimally at 1.7-2.5 M NaCl, with a minimum requirement of about 1.0 M NaCl (Mullakhanbhai & Larsen, 1975). Its internal KCl concentration is likely to fall into a similar range. The data show that citrate synthase is active at KCl and NaCl concentrations significantly below this range; approximately 50% of the activity observed at 2 M NaCl or KCl is retained at 0.1 M (Figure 3.1).

The activity profiles for the two cations are very similar up to a concentration of 1 M, at which point further increases in K^+ exert no change on activity, while increases in Na^+ give a proportional reduction in activity. The higher concentrations of Na^+ probably induce structural changes that affect activity through changes in the way the enzyme binds hydrated salt ions. The significantly greater strength of the hydration interactions of Na^+ over those of K^+ may contribute to this. The absence of high-salt inactivation in KCl solutions is not surprising, considering the presence within *Hf.volcanii* of high KCl concentrations *in vivo*.

At 25°C *Hf.volcanii* citrate synthase is remarkably stable at low salt concentrations, even considering the organism's moderate salt requirements. The enzyme remains stable at salt concentrations in which it is active, down to 0.1 M KCl. In contrast, the citrate synthase of the extreme halophile *Halobacterium cutirubrum* exhibits some activity at 0.1 M KCl but has a half-life of only 4 minutes at 1 M KCl (Higa & Cazzulo, 1975).

The inactivation temperatures were increased from 25°C until measurable inactivation rates were observed in 2 M KCl. The inactivation data, displayed as Arrhenius plots in Figures 3.6a and 3.6b, show that

Hf.volcanii citrate synthase is stable up to approximately 60°C. This value is of the order expected for an organism growing at up to 45°C. The rate of activity loss changes uniformly with the small variation in temperature, the slope of the graph giving the enthalpy change involved in formation of the transition state from the native enzyme during the inactivation process (980 kJ mol^{-1} in this case, values for pig heart, *Tp.acidophilum* and *S.solfataricus* being 331, 364 and 635 kJ mol^{-1} respectively). The displacement of the slope along the X-axis gives an indication of thermostability of the enzyme relative to those of *S.solfataricus*, *Tp.acidophilum* and pig heart.

The slope of the *Hf.volcanii* citrate synthase Arrhenius plot is significantly greater than those of the enzymes from pig heart or *Tp.acidophilum*. It is thought that the enthalpy change required to go from the native protein to the transition state is related to the breaking of electrostatic interactions, while the entropy change is related to the disruption of hydrophobic interactions. This suggests that heat denaturation of the *Hf.volcanii* enzyme requires more electrostatic interactions to be broken in reaching the transition state than in the others. It may be that surface and internal electrostatic interactions make a greater contribution to thermostability in the halphilic enzyme than in the pig heart or *Tp.acidophilum* enzymes..

At 2 M KCl and 66°C the rate constant of inactivation of the *Hf.volcanii* citrate synthase is 0.23 min^{-1}, while at 3 M KCl and 40°C the rate constant of inactivation of the *H.cutirubrum* citrate synthase is 0.35 min^{-1} (Higa & Cazzulo, 1975). Thus the *Hf.volcanii* enzyme is much more stable at high salt concentrations as well as at low salt concentrations.

Once measurable rates of inactivation had been achieved, the effect of KCl concentration on stability was investigated. The data in Figure 3.7 indicate that stability is highly dependent on KCl concentration between 1.6 and 3 M. Figure 3.8 displays the same data, relating half-life to salt concentration. It is possible that at high salt concentrations (> 3 M) *Hf.volcanii* citrate synthase is more thermostable than would be expected (i.e. remaining stable for periods of hours at temperatures more than 20°C above the parent organism's maximum growth temperature); however, this was not investigated. It is interesting to note that an increased requirement for NaCl has been observed when the growth temperature of *Hf.volcanii* is raised (Mullakhanbhai & Larsen, 1975).

Once *Hf.volcanii* citrate synthase is over-expressed in either a mesophilic or halophilic host, it will be necessary to compare its physico-chemical characteristics with those of the wild-type enzyme. This comparison must cover both its kinetic parameters and its salt requirements for stability and activity. Techniques such as differential scanning calorimetry, to determine ΔG of stability, coupled with circular dichroism or fluorescence to measure unfolding, are required for a detailed analysis of enzyme stability. The methods described above are limited in that they rely upon enzyme activity as a measure of unfolding; however, they could provide a simple means of investigating the salt requirements of over-expressed halophilic citrate synthases.

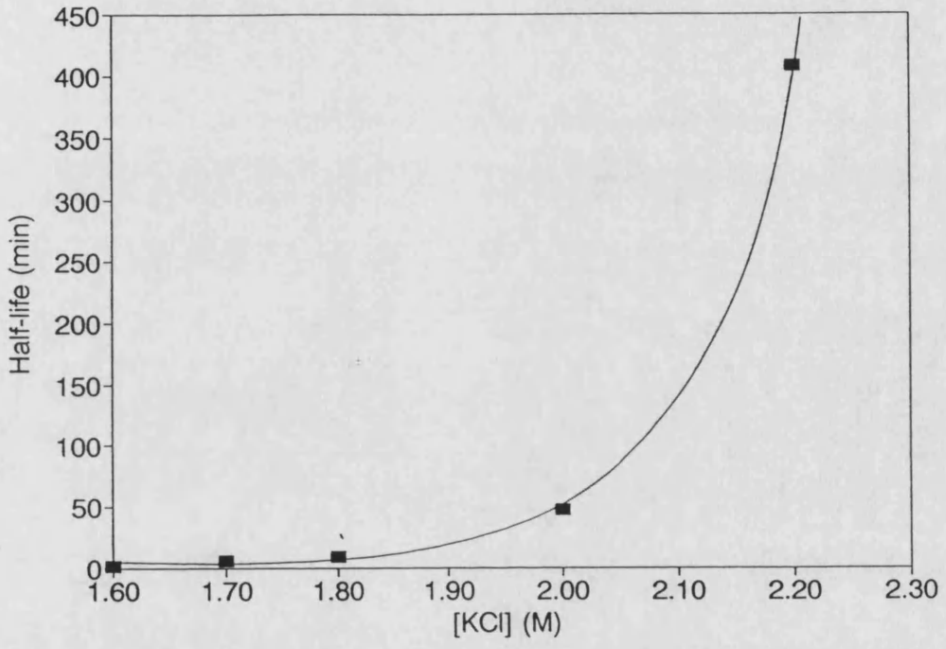

Figure 3.8 Effect of KCl concentration on the half-life of citrate synthase at 64°C

The graph shows half-life in miutes plotted against KCl concentration (Molar).

Chapter 4

Creation of size-restricted genomic libraries and screening for the citrate synthase gene

4.1 Introduction

Citrate synthase from *Hf.volcanii* was purified in our laboratory by Dr. M.J. Bonete, University of Alicante, Spain using a scheme based on that used to purify the *H.halobium* citrate synthase (James, 1990). SDS-PAGE of the purified enzyme revealed the presence of two major protein bands of apparent molecular weight 45 kD and 55 kD. These were Western blotted onto Immobilon P membrane and both bands were cut out and sequenced directly from the membrane.

Sequences were obtained for both bands. These are shown below, using single letter amino acid codes where X indicates an unknown residue and residues in parentheses are tentative assignments.

55 kD XXXXVIAAAYRTPFGKAA

45 kD (S)GELKRGLEGVLVAESKLSFIDGDAGQLVYXGYDIEDL

The 45 kD band appears to be citrate synthase as judged by a SWISSPROT database search and sequence alignment described in the Appendix. Outlines of the purification scheme, blotting and sequencing are also given in the Appendix.

From the N-terminal peptide sequence of the protein, oligonucleotide probes were designed and used to screen Southern blots of restriction digested genomic DNA. A suitable size range of positively-hybridising fragments were identified and used to create a size-restricted genomic library in the pUC18 plasmid vector. This was in turn screened by hybridisation and positive clones were further analysed by restriction digests and sequencing of the plasmid DNA.

4.2 Oligonucletide design

Oligonucleotides were designed from the N-terminal sequence with a bias towards G and C at the third base of each codon, reflecting the high G+C content of the *Hf.volcanii* genome. In order to aid selection of third bases a composite codon usage table was compiled from the dihydrolipoamide dehyrogenase (Vettakkorumakankav *et al.,* 1992), dihydrofolate reductase (Zusman *et al.,* 1989), his C (Conover & Doolittle, 1990) and trp cluster (Lam *et al,.* 1990) genes of *Hf.volcanii.* Figure 4.1 summarises this information.

4.3 Southern blot analysis

Initially 10 µg aliquots of genomic DNA were digested with single restriction enzymes. Once restriction fragments were separated by agarose gel electrophoresis and immobilised on nylon membranes by Southern transfer, hybridisation to a γ ^{32}P end-labelled oligonucleotide was carried out. The sizes of hybridising fragments were calculated from autoradiographs of the blots.

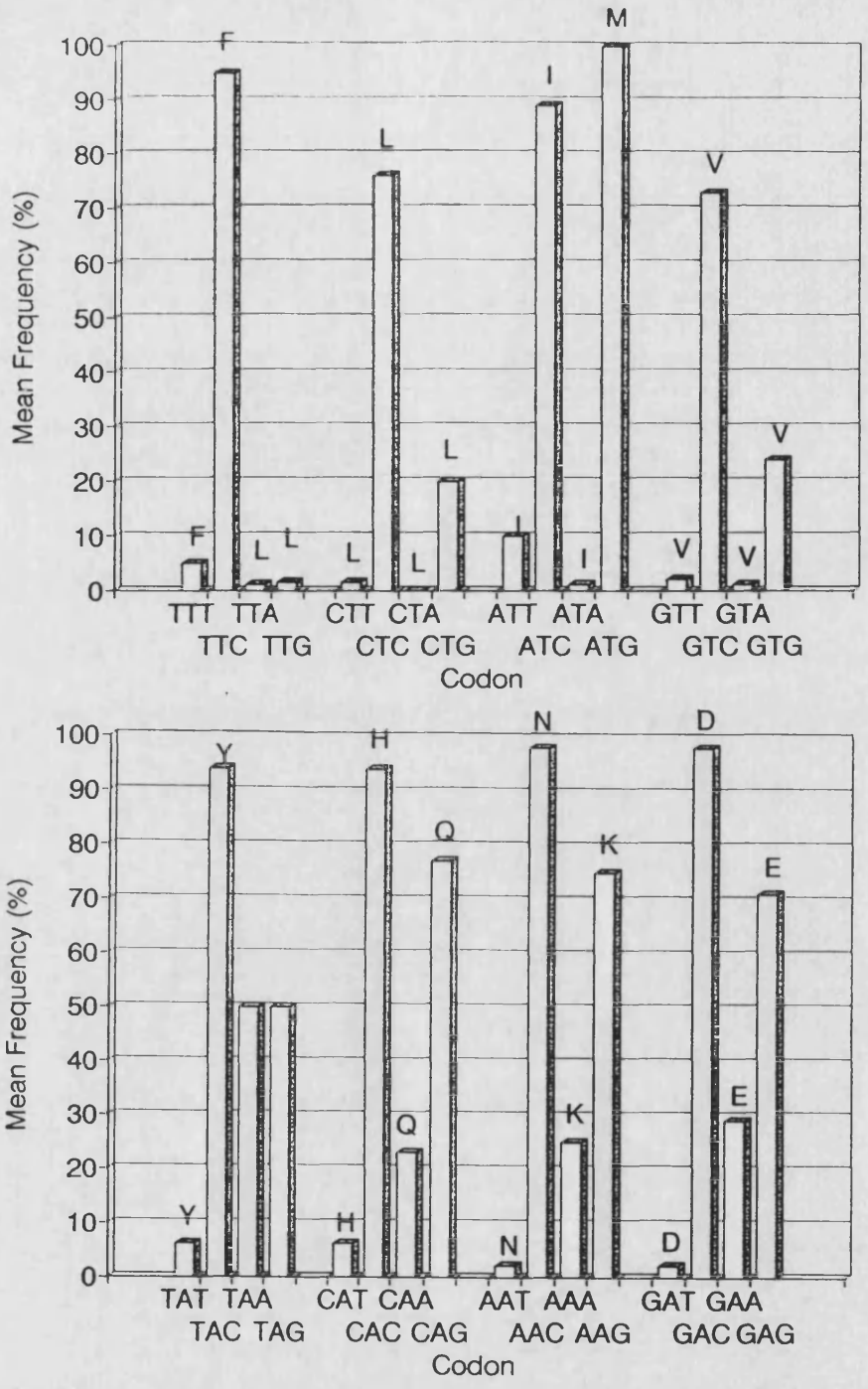

Figure 4.1 Codon preference in *Hf.volcanii*

The frequency of each codon is expressed as a percentage of the total number coding for a particular amino acid.

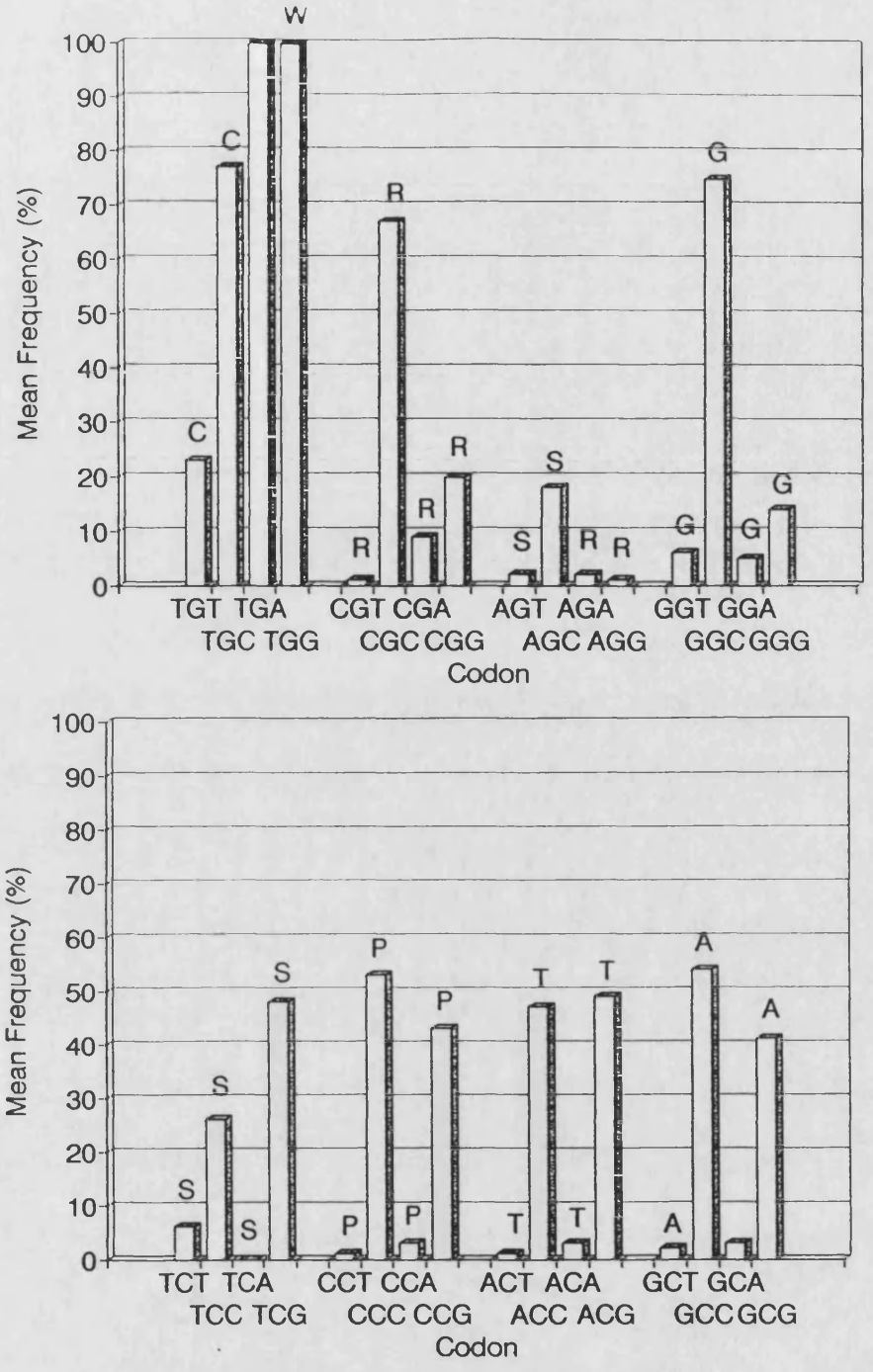

Figure 4.1 Codon preference in *Hf.volcanii* (continued)

The process was repeated for genomic DNA digested with pairs of restriction enzymes in order to identify hybridising fragments smaller than those produced by single enzymes ie. fragments that were doubly digested.

4.4 Creation of a size-restricted genomic library

Approximately 150 µg genomic DNA was digested with the required pair of restriction enzymes and the fragments were separated on a 1% agarose gel. A gel slice containing fragments of approximately the desired size was cut and the DNA recovered by freeze-squeeze. The fragments were purified using a NENsorb cartridge.

A 100 ng aliquot of commercially prepared pUC18 *Eco*RI / BAP plasmid (ie. digested with *Eco*RI and treated with alkaline phosphatase) was digested with *Pst*I and purified using a NENsorb cartridge. Ligations were set up using 10 ng plasmid DNA with 10%, 1% and 0.1% of the purified fragments, each in a volume of 20 µl.

Competent *E.coli* JM103 were prepared by the calcium chloride method and a 200 µl volume transformed with 8 µl ligated DNA. Transformants were selected by plating out on 2xYT agar containing ampicillin (100 µg ml^{-1}).

4.5 Screening for positive clones

Transformants were streaked onto gridded 2xYT agar plates containing ampicillin (100 µg ml^{-1}). From these master plates transformants were streaked onto gridded nylon membranes which were treated as described in section 2.3.9.3 Colony Hybridisation. Positive clones were purified by plating out and screening single colonies by

streaking onto gridded membranes and hybridising as above. This was repeated until all the colonies screened gave a positive signal. Clones were stored as glycerol stocks at -20°C.

4.6 Results

4.6.1 Screening Southern blots with a non-degenerate 44 base oligonucleotide probe

Figure 4.2 shows the sequence of a 44 base oligonucleotide designed from the citrate synthase N-terminal sequence. Its estimated T_m is in excess of 100°C (using the formula T_m = 4(Number of G and C bases) + 2(Number of A and T bases)).

When used to probe a Southern blot of the restriction digests shown in Figure 4.3a, a *Pst*I fragment of approximately 7.2 kb and a *Pst*I / *Eco*RI fragment of approximately 2.9 kb were identified as shown in Figure 4.3b (lanes 2 and 5 respectively). Hybridisations and washes were carried out at 60°C.

A library of *Pst*I / *Eco*RI fragments in the 3 kb size range were cloned into pUC18 as described in Section 5.4. Ligation of inserts into the cut vector yielded 534 transformants. Control transformations with ligations containing vector plus insert (no ligase), vector only (plus ligase) and insert only (plus ligase) yielded a total of 4 transformants between them. A control transformation with 1ng of uncut pUC18 gave an estimated transformation efficiency of 1.5×10^7 μg^{-1}.

```
G E L K R G L E G V L  V   A   E   S   K   L   S   F   I   D   G   D   A   G   Q  L V Y X G Y D I E D L

                      GTA GCA GAA TCA AAA CTA TCA TTT ATA GAT GGA GAT GCA GGA CAA
                        G   G   G   G   G   G   G   C   C   C   G   C   G   G   G
                        C   C       C       C   C       T       C       C   C
                        T   T       T       T   T                   T       T   T
                                  AGC     TTA AGC
                                    T       G   T

                      GTC GCC GAC TCG AAG CTT TCG TTC ATC GAC GGC GAC GCC GGC CA
```

Figure 4.2 Sequence of 44 base oligonucleotide probe

The citrate synthase N-terminal sequence is shown, with all possible codons below each residue. The probe sequence is shown in **bold**.

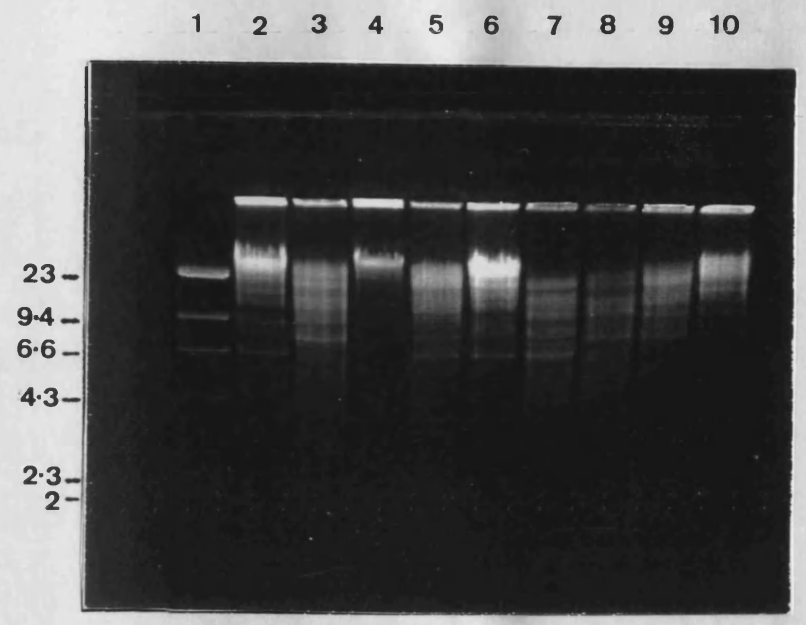

Figure 4.3a Electrophoresis of genomic DNA restriction digests on 0.6% agarose gel

Lane 1: λ *Hin*dIII
Lane 2: *Pst*I
Lane 3: *Sma*I
Lane 4: *Bam*HI
Lane 5: *Pst*I / *Eco*RI
Lane 6: *Pst*I / *Bam*HI
Lane 7: *Pst*I / *Sma*I
Lane 8: *Sma*I / *Eco*RI
Lane 9: *Sma*I / *Bam*HI
Lane 10: *Bam*HI / *Eco*RI

Fragment sizes are indicated in kilobases. Figure 4.3b shows an autoradiograph of the corresponding Southern blot.

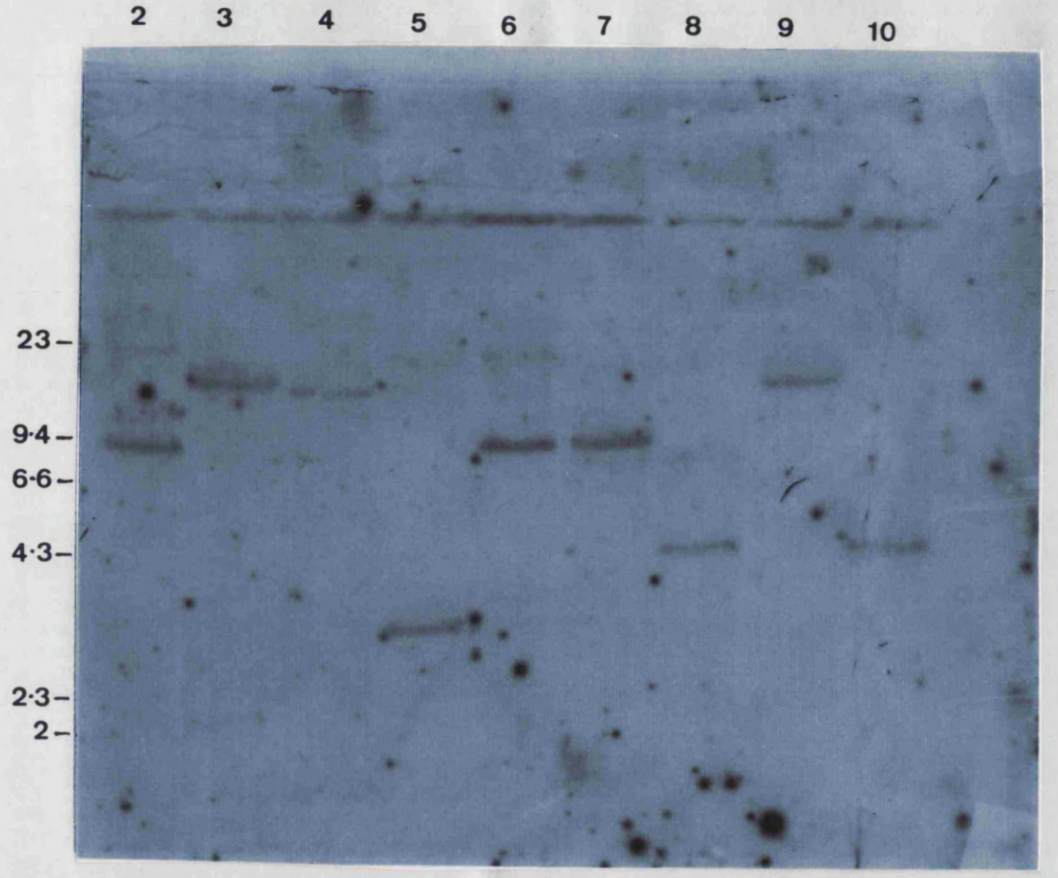

Figure 4.3b Autoradiograph of Southern blot probed with non-degenerate 44 base oligonucleotide (4 day exposure)

Lane 1: Not shown
Lane 2: *Pst*I
Lane 3: *Sma*I
Lane 4: *Bam*HI
Lane 5: *Pst*I / *Eco*RI
Lane 6: *Pst*I / *Bam*HI
Lane 7: *Pst*I / *Sma*I
Lane 8: *Sma*I / *Eco*RI
Lane 9: *Sma*I / *Bam*HI
Lane 10: *Bam*HI / *Eco*RI

The blot was washed at 60°C. Fragment sizes are indicated in kilobases. Figure 4.3a shows the corresponding agarose gel.

The transformants were screened by colony hybridisation as described above, using the same 44 base oligonucleotide. Hybridisations and washes were carried out at 60°C. Four strongly hybridising clones were indentified, plus ten that hybridised less strongly. Examples of each are shown in Figure 4.4.

All twelve positive clones were purified as described in Section 5.5, using the same hybridisation and washing conditions as the initial screening. After the first round of purification it became apparent that only the four strongly hybridising clones were true positives, as the others did not demonstrate reproducible hybridisation. Plasmid DNA was prepared from the four positive clones by the caesium chloride 'Maxiprep' method.

4.6.2 Restriction and hybridisation analysis of positive clones

Initially a dot-blot of 1 µg DNA from each positive clone was prepared, together with 1 µg pUC18 DNA as a control. This was probed with the 44 base oligonucleotide, hybridising and washing at 60°C in 2xSSC. Hybridisation to the plasmids containing the insert was observed, while there was no detectable hybridisation to pUC18 alone. Restriction digests were carried out using the enzymes *Pst*I, *Eco*RI, *Hin*dIII and *Ava*I. All of the positive clones appeared to contain the same insert, as judged by the restriction pattern.

Figure 4.5a shows the restriction fragments of one positive clone separated by agarose gel electrophoresis. These were transferred to nylon membrane by Southern blotting and probed with the 44 base oligonucleotide (Figure 4.5b).

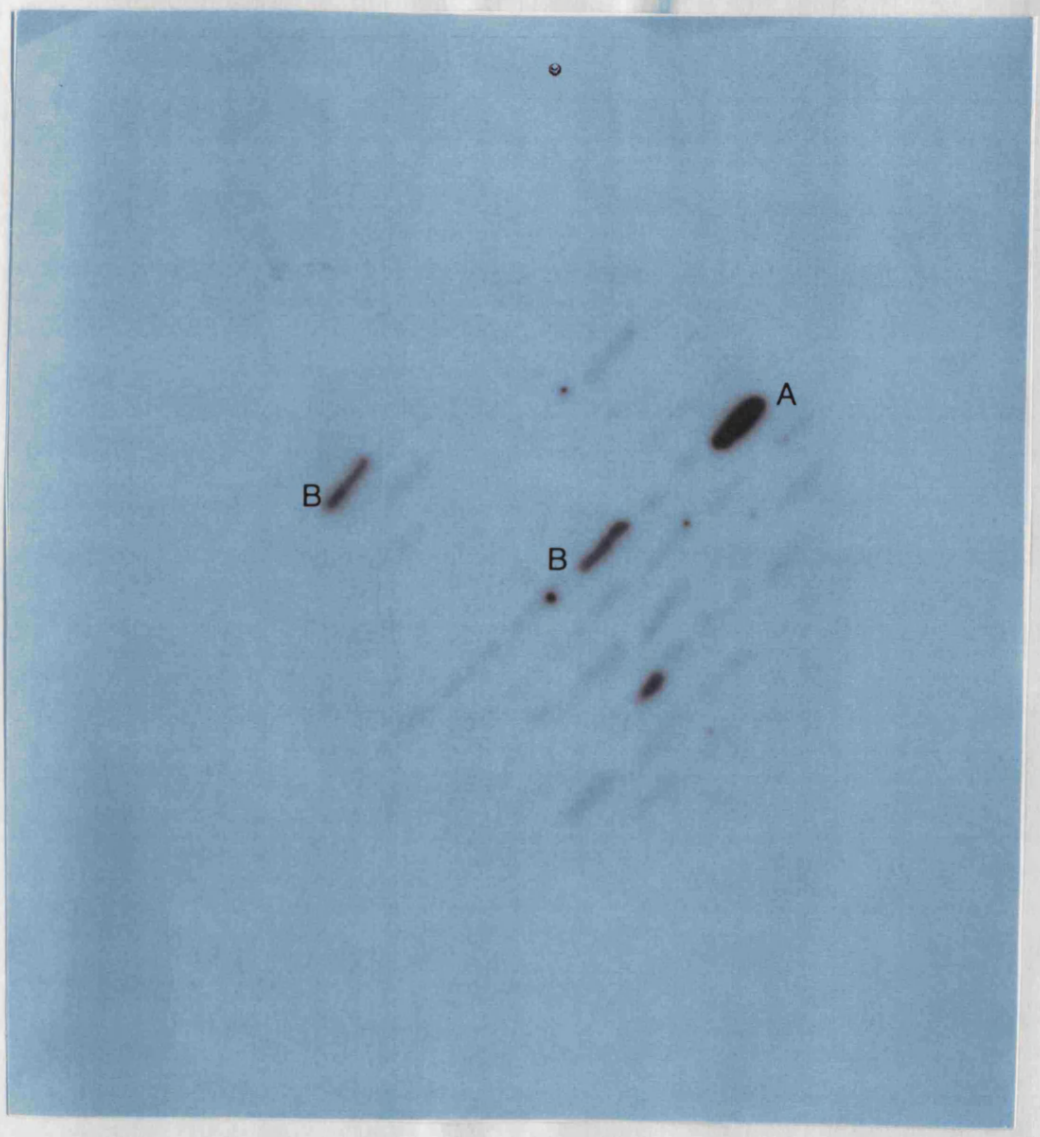

Figure 4.4 Autoradiograph of colony blot probed with 44 base non-degenerate oligonucleotide (2 day exposure): screening transformants

A: Strong hybridisation (true positive)
B: Weaker hybridisation (later identified as false positive)

The blot was washed at 60°C.

Further restriction digests were carried out the enzymes *Acc*I, *Ava*I and *Hin*dIII. In Figure 4.6a the restriction fragments are shown, separated by agarose gel electrophoresis. Figure 4.6b shows a Southern blot of those fragments, probed with the 44 base oligonucleotide.

From these data an initial restriction map was drawn (Figure 4.7), although the exact binding site of the probe cannot be deduced. Fragments of the 2.9 kb insert were therefore sub-cloned into pUC18 for a more detailed analysis.

4.6.3 Sub-cloning fragments from the 2.9kb positive clone

Aliquots of 5 µg plasmid DNA were digested with *Ava*I and *Hin*dIII. A 1.5 kb *Ava*I fragment, 1.5 kb *Hin*dIII fragment and 1.2 kb *Hin*dIII fragment were gel purified and cleaned by passing through a NENsorb cartridge. Vector was prepared by digesting 1 µg aliquots of pUC18 DNA with either *Ava*I or *Hin*dIII, treating with alkaline phosphatase and cleaning using a NENsorb cartridge. Each of the fragments was ligated into appropriately cut vector.

Sub-clones of the two *Hin*dIII fragments were obtained. The *Ava*I fragment proved difficult to clone, and despite several attempts no transformants were obtained. Since the only sub-fragment of the 2.9 kb clone to retain full hybridisation is that produced by *Ava*I digestion, cloning and sequencing this fragment is particularly important with regard to locating the hybridisation site of the probe.

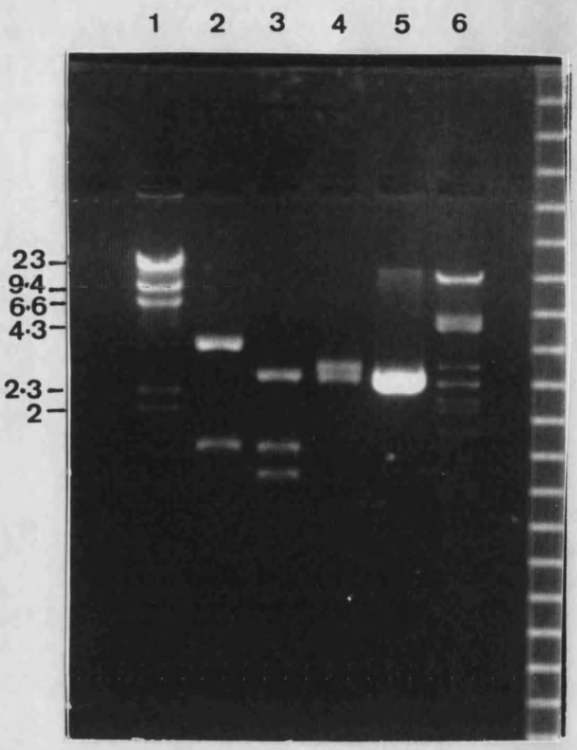

Figure 4.5a Electrophoresis of restriction digests of 2.9 kb *Pst*I / *Eco*RI clone on 1% agarose gel

Lane 1: λ *Hin*dIII		
Lane 2: *Ava*I	Fragment sizes:	(3.7), 1.5 kb
Lane 3: *Hin*dIII		(2.7), 1.5, 1.2 kb
Lane 4: *Pst*I / *Eco*RI		2.9, (2.7) kb
Lane 5: pUC18 *Eco*RI		(2.7) kb
Lane 6: λ *Pst*I		

Restriction fragment sizes are indicated in kilobases (the fragment containing the pUC18 vector in parentheses).

Figure 4.5b shows the corresponding autoradiograph.

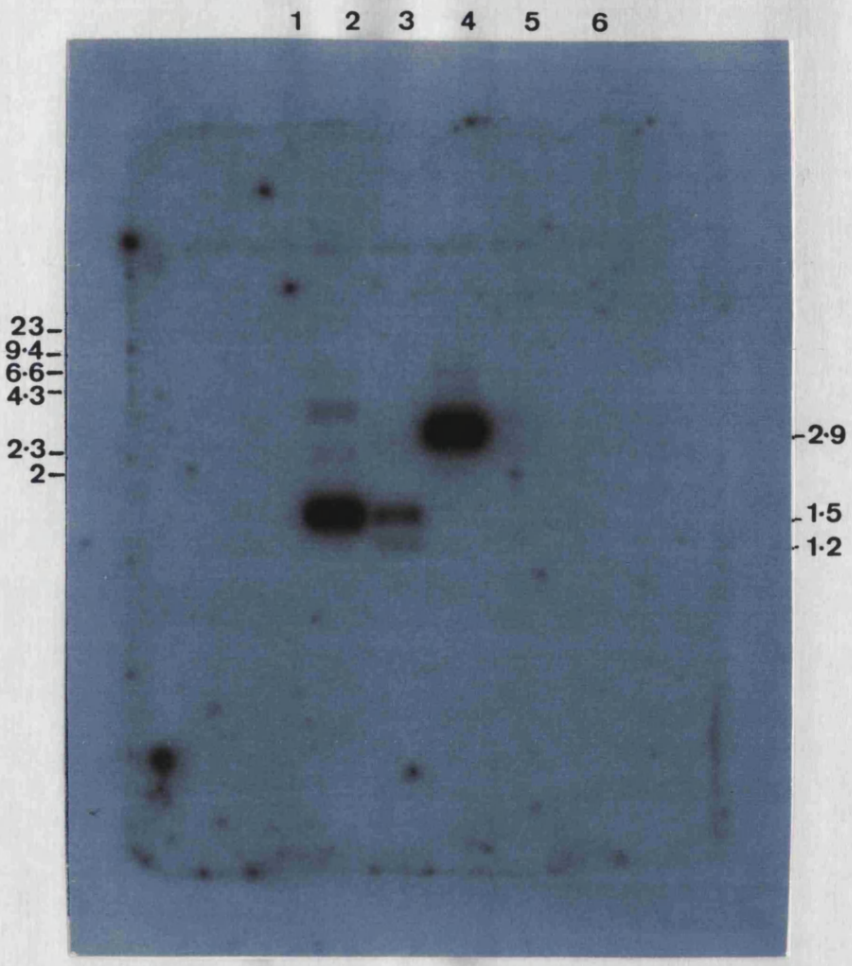

Figure 4.5b Autoradiograph of Southern blot probed with 44 base non-degenerate oligonucleotide (4 day exposure)

Lane 1: λ *Hin*dIII
Lane 2: *Ava*I
Lane 3: *Hin*dIII
Lane 4: *Pst*I / *Eco*RI
Lane 5: pUC18 *Eco*RI
Lane 6: λ *Pst*I

Restriction fragment sizes are indicated in kilobases. Figure 4.5a shows the corresponding agarose gel. The blot was washed at 60°C.

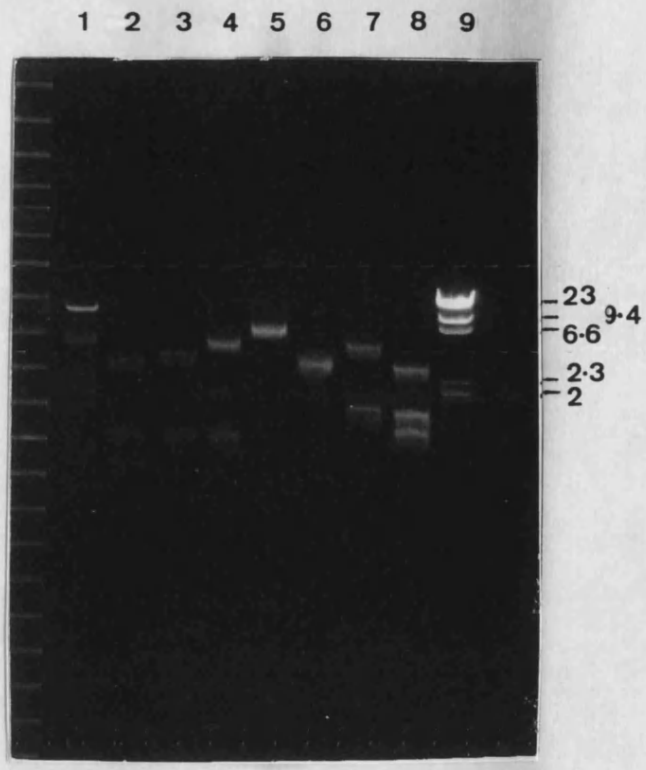

Figure 4.6a Electrophoresis of restriction digests of 2.9 kb *Pst*I / *Eco*RI clone on 1% agarose gel

Lane 1: λ *Pst*I		
Lane 2: *Ava*I / *Hin*dIII	Fragment sizes:	(2.7), 1.2, 0.95 kb
Lane 3: *Ava*I / *Acc*I		(3.3), 1.2, 0.7 kb
Lane 4: *Acc*I		(4.0), 1.2 kb
Lane 5: *Eco*RI		(5.7) kb
Lane 6: *Pst*I / *Eco*RI		(2.8) kb
Lane 7: *Ava*I		(3.8), 1.5 kb
Lane 8: *Hin*dIII		(2.7), 1.5, 1.2 kb
Lane 9: λ *Hin*dIII		

Restriction fragment sizes are indicated in kilobases (the fragment containing the pUC18 vector in parentheses).Figure 4.6b shows the corresponding autoradiograph.

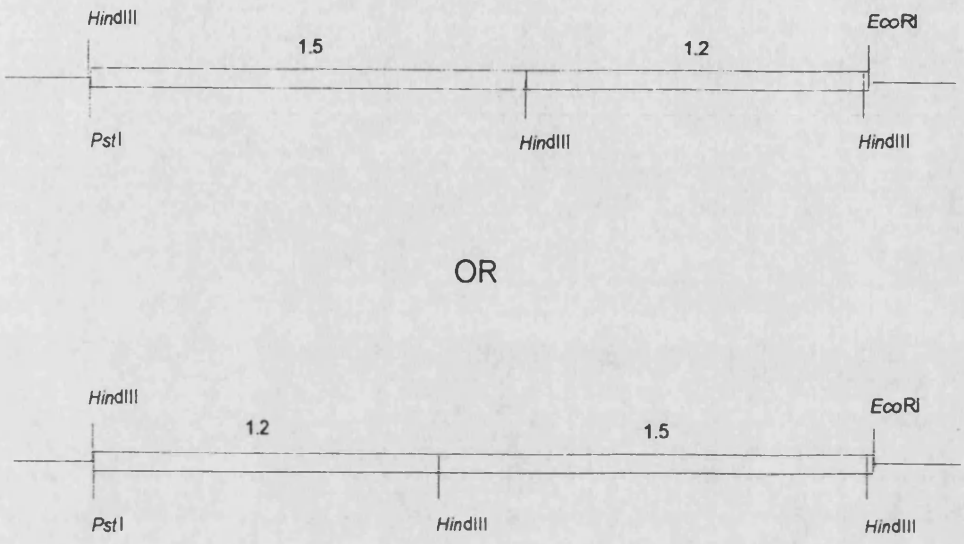

Figure 4.7 Summary of initial restriction site information obtained from 2.9 kb *Pst*I / *Eco*RI clone in pUC18

No sites within the clone could be mapped unambiguously at this stage. In addition to the *Hin*dIII sites are: a pair of *Ava*I sites 1.5 kb apart and a pair of *Acc*I sites 1.2 kb apart. There are no *Ava*I or *Acc*I sites in the vector.

4.6.4 Restriction analysis of sub-cloned *Hind*III fragments

Both the cloned 1.5 kb and 1.2 kb *Hind*III fragments were digested with *Hind*III, *Ava*I and *Acc*I. Figures 4.8 and 4.9 show the restriction fragments separated by agarose gel electrophoresis.

4.6.5 Sequence analysis of 2.9 kb *Pst*I / *Eco*RI clone and *Hind*III sub-clones

In order to orientate the two *Hind*III sub-clones and to make an initial search for citrate synthase sequence, the ends of the 2.9 kb *Pst*I / *Eco*RI clone and *Hind*III subclones were sequenced, initially using the Forward and Reverse universal primers of pUC18 and then using synthetic 17 base primers to extend the sequenced region. The locations of the regions sequenced are shown in Figure 4.10, which also indicates the deduced orientations of the subclones.

The sequences determined were translated in all reading frames (including those for the complementary strand) and used to search the Swissprot protein database via the FASTA program from the GCG package. The translated sequences did not show any significant identity with citrate synthase, although a putative open reading frame of an unrelated gene was identified. This open reading frame was sequenced and is described in more detail in Chapter 7.

Figure 4.8 Electrophoresis of restriction digests of 2.9 kb *Pst*I / *Eco*RI clone and *Hind*III subclones on 1% agarose gel

Lane 1: 1.5 kb *Hind*III subclone: *Hind*III digest	(2.7), 1.5 kb
Lane 2: 1.5 kb *Hind*II subclone: *Hind*III / *Ava*I digest	(2.7), 0.95, 0.3 kb
Lane 3: 1.2 kb *Hind*III subclone: *Hind*III digest	(2.7), 1.2 kb
Lane 4: 1.2 kb *Hind*III subclone: *Hind*III / *Ava*I digest	(2.7), 1.2 kb
Lane 5: 2.9 kb *Pst*I / *Eco*RI clone: *Hind*III digest	(2.7), 1.5, 1.2 kb
Lane 6: 2.9 kb *Pst*I / *Eco*RI clone: *Hind*III / *Ava*I digest	(2.7), 1.2, 0.95, 0.3 kb
Lane 7: λ *Pst*I	

Restriction fragment sizes are indicated in kilobases (the fragment containing the pUC18 vector in parentheses).

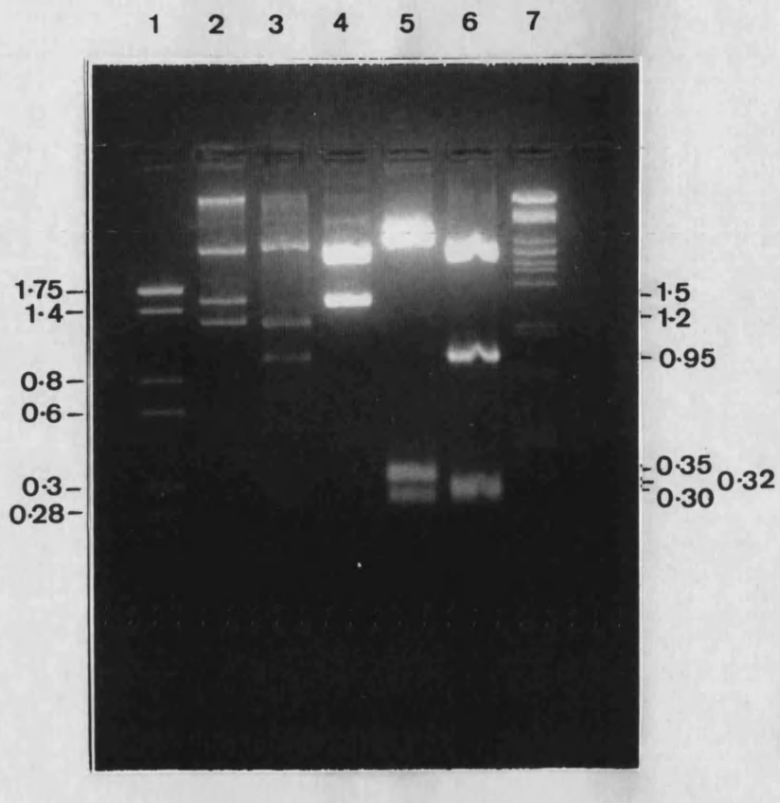

Figure 4.9 Electrophoresis of restriction digests of 2.9 kb *Pst*I / *Eco*RI clone and 1.5 kb *Hin*dIII subclone on 2% agarose gel

Lane 1: Low range DNA markers	
Lane 2: 2.9 kb *Pst*I / *Eco*RI clone: *Hin*dIII digest	(2.7), 1.5, 1.2 kb
Lane 3: 2.9 kb *Pst*I / *Eco*RI clone: *Hin*dIII / *Ava*I digest	(2.7), 1.2, 0.95, 0.32 kb
Lane 4: 1.5 kb *Hin*dIII subclone: *Hin*dIII digest	(2.7), 1.5 kb
Lane 5: 1.5 kb *Hin*dIII subclone: *Ava*I digest	(3.3), 0.35, 0.3
Lane 6: 1.5 kb *Hin*dIII subclone: *Hin*dIII / *Ava*I digest	(2.7), 0.95, 0.32
Lane 7: λ *Pst*I	

Restriction fragment sizes are indicated in kilobases (the fragment containing the pUC18 vector in parentheses).

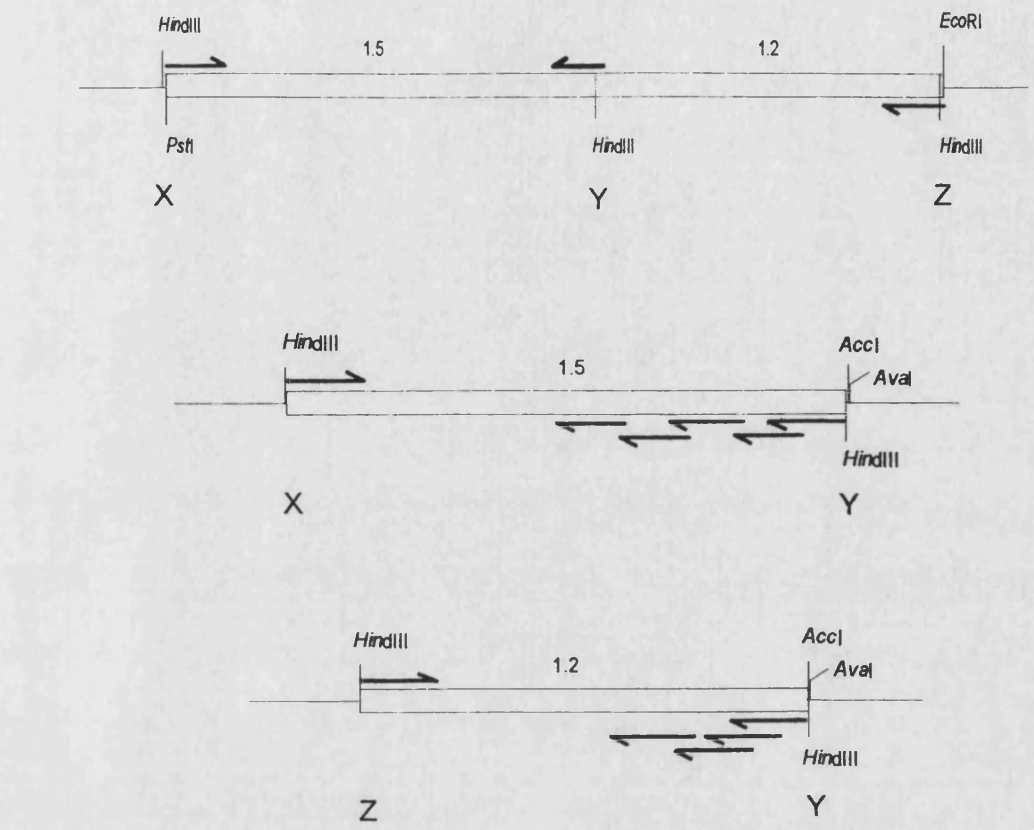

Figure 4.10 Regions of 2.9 kb *Pst*I / *Eco*RI clone and *Hind*III subclone sequenced.

Distances are indicated in kilobases. X, Y and Z indicate the orientation of the *Hind*III fragments.

The sequence data (including that presented in Chapter 7) are combined with the results of the restriction digests described above to give the restriction maps shown in Figure 4.11.

4.7 Screening with a pair of degenerate 17 base oligonucleotide probes

As a unique 44 base oligonucleotide had failed to identify a citrate synthase gene, a different approach was taken in designing new oligonucleotides from the citrate synthase N-terminal sequence. Furthermore, as the long oligonucleotide appeared to bind to an unrelated gene, it was necessary to consider the possibilty that other oligonucleotides designed to the same stretch of amino acid sequence also bind to that gene.

A pair of non-overlapping, degenerate 17 base oligonucleotides (referred to here as A and B) were designed to the citrate synthase N-terminal sequence as shown in Figure 4.12. Their T_ms were estimated as 46-56°C for oligonucleotide A and 40-52°C for oligonucleotide B, using the formula described for the 44 base probe.

These were used as probes with the aim of identifying a restriction fragment to which both would bind independently. Ideally the two non-overlapping regions of sequence used would not include the region of sequence used to design the previous 44 base probe, in order to minimise the risk of hybridisation to the unrelated gene described in Chapter 7.

In a degenerate oligonucleotide mixture the number of different species must be kept below the limit where the "correct" sequence is present in such a small proportion as to be undetectable. It was therefore

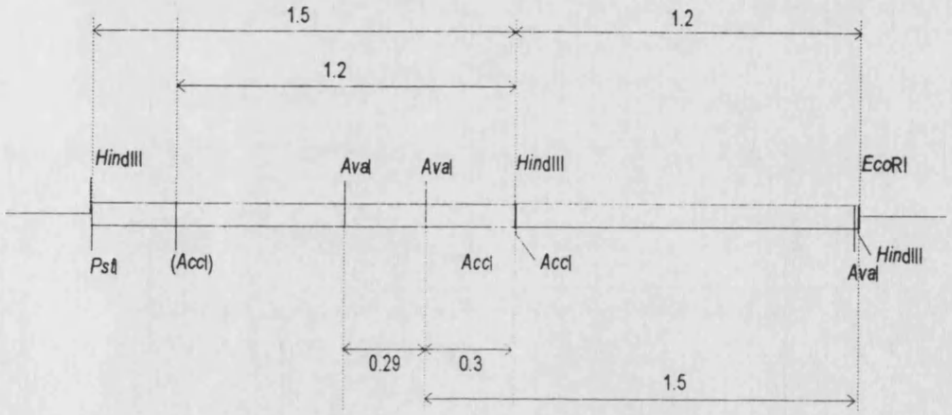

Figure 4.11 Restriction map of 2.9 kb PstI / EcroRI clone

This map was compiled from restriction digest and sequence data. The restriction sites in parentheses were not confirmed by sequencing. Distances are indicated in kilobases.

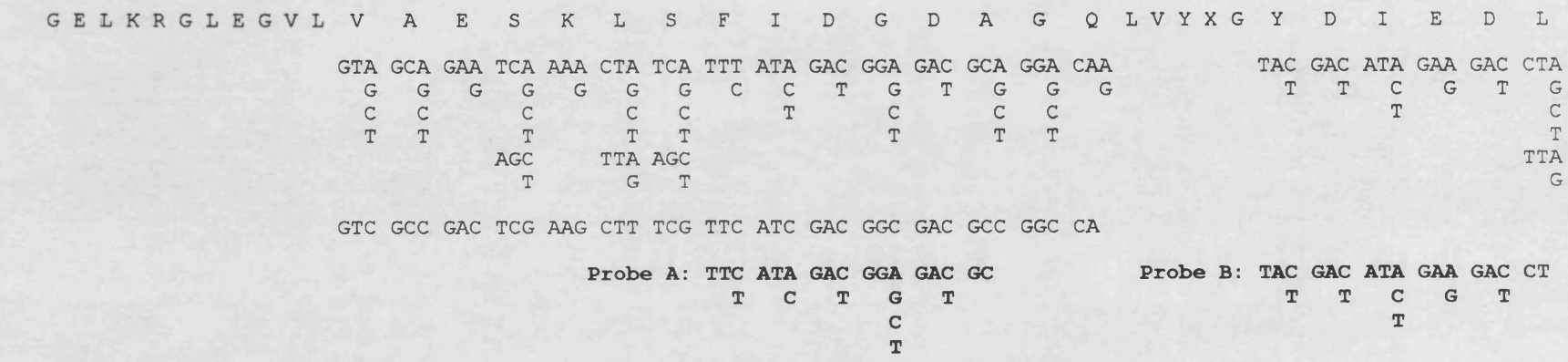

Figure 4.12 Sequence of 17 base oligonucleotide probes

The citrate synthase N-terminal sequence is shown, with all possible codons below each residue. The probe sequences are shown in **bold**, including degeneracies. The 44 base probe is also shown.

decided to limit the number of possible sequences for each oligonucleotide to a maximum of 96. To avoid exceeding this limit, one of the two oligonucleotides (oligo A) had to be made to a region overlapping that used to design the 44 base probe.

4.7.1 Preliminary dot-blot results

Initial hybridisations were carried out with both 17mers to dot-blots of 1 µg of the 2.9 kb *Pst*I / *Eco*RI clone and 1 µg genomic DNA to confirm that they hybridised to genomic DNA and not to the previously-cloned open reading frame. Both probes were found to hybridise to genomic DNA at 30°C, while neither hybridised to the 2.9 kb *Pst*I / *Eco*RI clone at this temperature. Oligonucleotide A hybridised more strongly than B, as judged by the intensity of the dots. In both cases hybridisation was poor, dots only becoming visible after a 14-day exposure.

4.7.2 Southern blot results

Hybridisation of oligonucleotide B was carried out at 30°C to Southern blots of genomic DNA digested with single restriction enzymes. After washing the blots at 30°C, autoradiographs showed that this probe was not hybridising under these conditions. This probe was not considered suitable for further investigation.

Oligonucleotide A was hybridised to Southern blots of genomic DNA digested with both single restriction enzymes and with pairs of enzymes. Initial blots were hybridised and washed at 37°C. Figures 4.13a and b show an example set of restriction digests and corresponding Southern

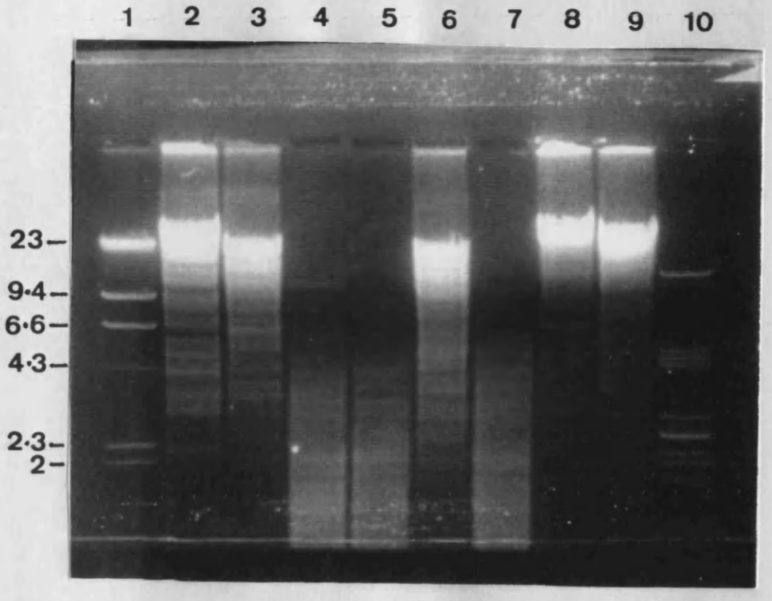

Figure 4.13a Electrophoresis of genomic DNA restriction digests on 0.6% agarose gel

Lane 1: λ *Hin*dIII
Lane 2: *Hin*dIII / *Pst*I
Lane 3: *Hin*dIII / *Sph*I
Lane 4: *Hin*dIII / *Sal*I
Lane 5: *Pst*I / *Sal*I
Lane 6: *Sph*I / *Pst*I
Lane 7: *Sph*I / *Sal*I
Lane 8: *Hin*dIII
Lane 9: *Sph*I
Lane 10: λ *Pst*I

Fragment sizes are indicated in kilobases. Figure 4.13b shows an autoradiograph of the corresponding Southern blot.

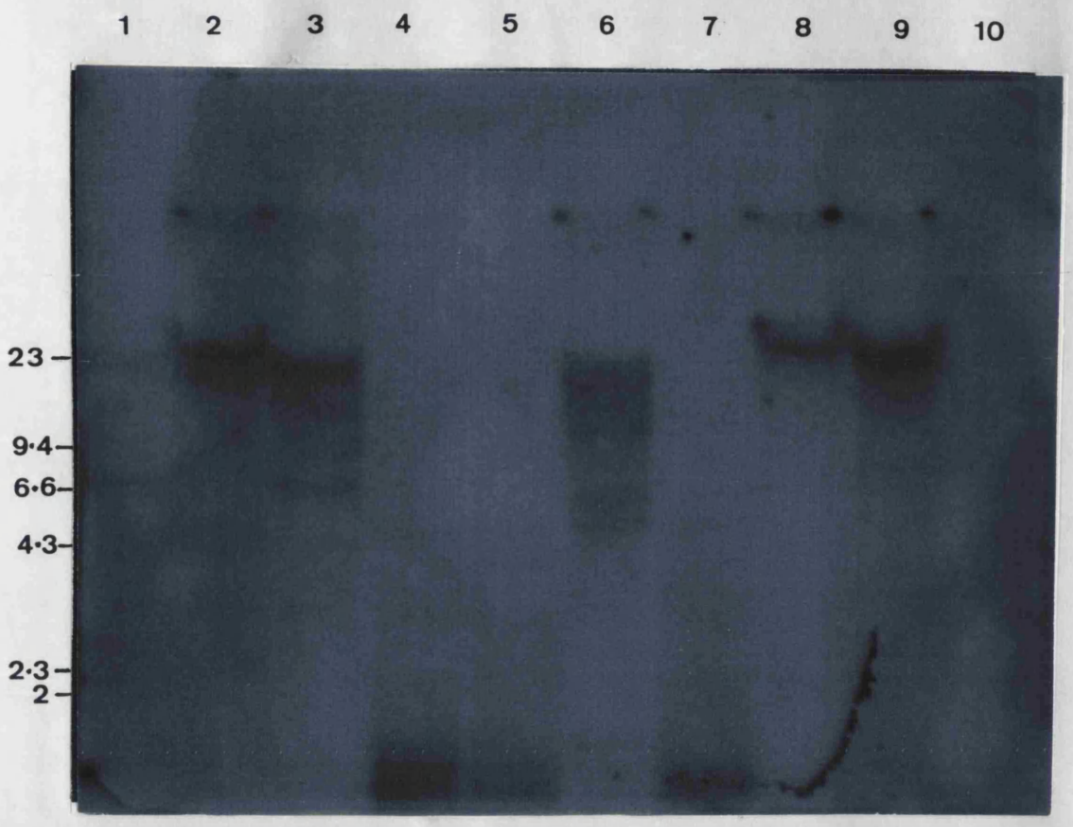

Figure 4.13b Autoradiograph of Southern blot probed with degenerate 17 base oligonucleotide A (9 day exposure)

Lane 1: λ *Hin*dIII
Lane 2: *Hin*dIII / *Pst*I
Lane 3: *Hin*dIII / *Sph*I
Lane 4: *Hin*dIII / *Sal*I
Lane 5: *Pst*I / *Sal*I
Lane 6: *Sph*I / *Pst*I
Lane 7: *Sph*I / *Sal*I
Lane 8: *Hin*dIII
Lane 9: *Sph*I
Lane 10: λ *Pst*I

Fragment sizes are indicated in kilobases. Figure 4.13a shows the corresponding agarose gel. The blot was washed at 37°C, in 2x SSC.

blot; hybridisation to a number of bands in each lane is apparent (autoradiograph exposed for 9 days).

As only one of the pair of probes (oligonucleotide A) exhibited potentially useful hybridisation, the possibility of two independent identifications of a restriction fragment containing the citrate synthase gene N-terminus was no longer viable. This meant that either such a fragment would have to be unambiguously indentified from the hybridisation of the remaining 17mer (ie. single bands would have to be identified on a Southern blot), or a new probe would have to be designed.

Before taking the step of involving new probes, the possibility of identifying single restriction fragments using oligonucleotide A alone was investigated.

In order to eliminate non-specific hybridisation and to obtain single bands in each lane, further blots were hybridised and washed at 42°C. Figures 4.14a and 4.14b show the result obtained under these conditions (autoradiograph exposed for 17 days). As multiple bands were still visible under these conditions, the stringencies of the hybridisation and wash were increased by raising the temperature of both to 45°C. This resulted in a further decrease in the intensity of all bands visible on the previous autoradiographs, without enhancement of any particular band.

In another approach to improving specificity, hybridisations were carried out at 37°C, where the oligonucleotide was known to bind efficiently, followed by washes where salt concentrations had been reduced to increase the stringency. These washes were at 37°C, in the same solution as describe in Section 2.3.9.1, except that the concentration of SSC was reduced from 2x to 0.5x. Figures 4.15a and 4.15b show that

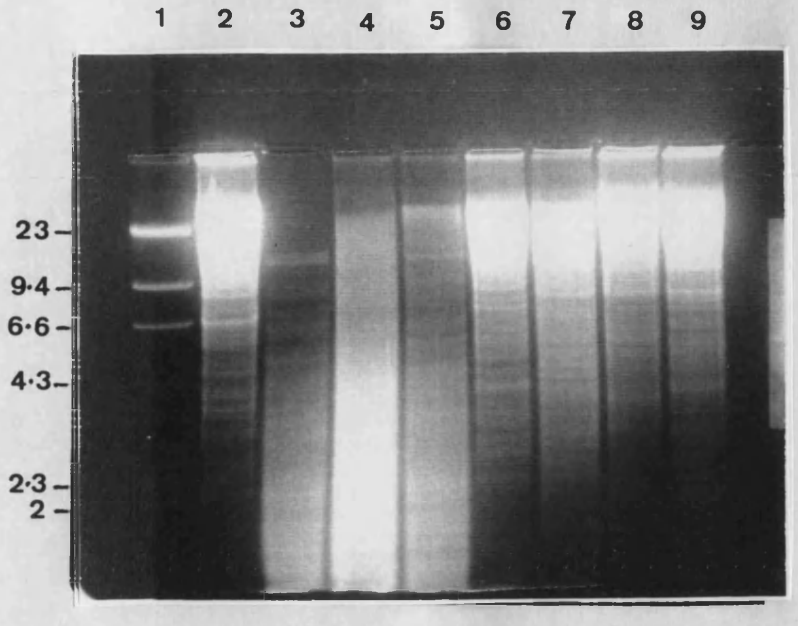

Figure 4.14a Electrophoresis of genomic DNA restriction digests on 0.6% agarose gel

Lane 1: λ *Hin*dIII
Lane 2: *Hin*dIII / *Sph*I
Lane 3: *Hin*dIII / *Sal*I
Lane 4: *Pst*I / *Sal*I
Lane 5: *Sph*I / *Sal*I
Lane 6: *Sph*I / *Pst*I
Lane 7: *Sma*I / *Eco*RI
Lane 8: *Sma*I / *Bam*HI
Lane 9: *Bam*HI / *Eco*RI

Fragment sizes are indicated in kilobases. Figure 4.14b shows an autoradiograph of the corresponding Southern blot.

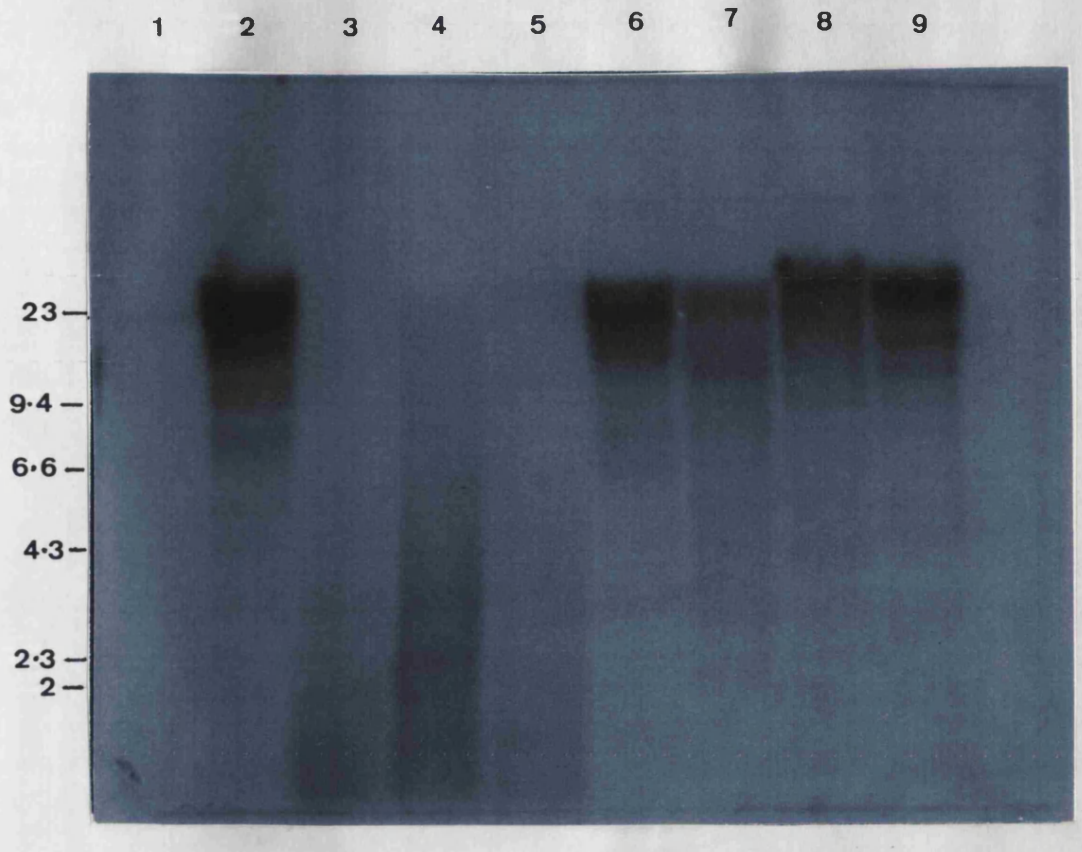

Figure 4.14b Autoradiograph of Southern blot probed with degenerate 17 base oligonucleotide A (17 day exposure)

Lane 1: λ *Hin*dIII
Lane 2: *Hin*dIII / *Sph*I
Lane 3: *Hin*dIII / *Sal*I
Lane 4: *Pst*I / *Sal*I
Lane 5: *Sph*I / *Sal*I
Lane 6: *Sph*I / *Pst*I
Lane 7: *Sma*I / *Eco*RI
Lane 8: *Sma*I / *Bam*HI
Lane 9: *Bam*HI / *Eco*RI

Fragment sizes are indicated in kilobases. Figure 4.14a shows the corresponding agarose gel. The blot was washed at 42°C, in 2x SSC.

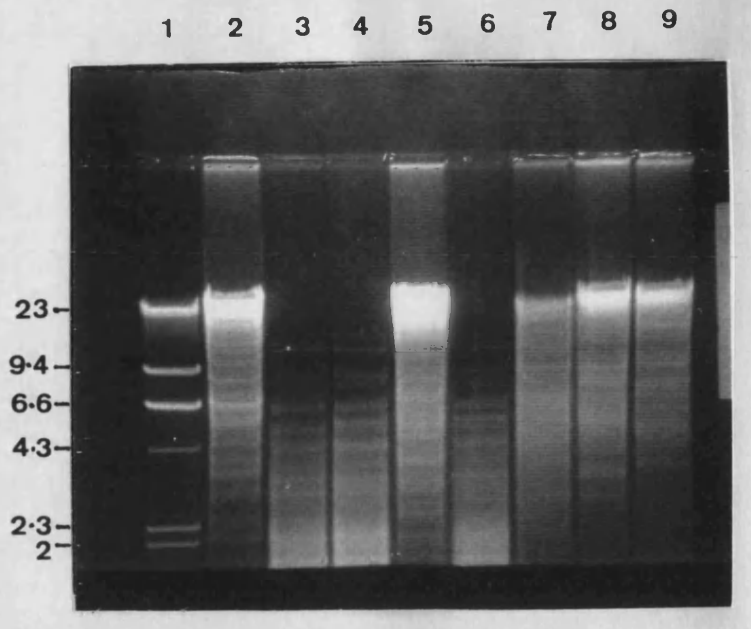

Figure 4.15a Electrophoresis of genomic DNA restriction digests on 0.6% agarose gel

Lane 1: λ *Hin*dIII
Lane 2: *Hin*dIII / *Sph*I
Lane 3: *Hin*dIII / *Sal*I
Lane 4: *Pst*I / *Sal*I
Lane 5: *Sph*I / *Pst*I
Lane 6: *Sph*I / *Sal*I
Lane 7: *Sma*I / *Eco*RI
Lane 8: *Sma*I / *Bam*HI
Lane 9: *Bam*HI / *Eco*RI

Fragment sizes are indicated in kilobases. Figure 4.15b shows an autoradiograph of the corresponding Southern blot.

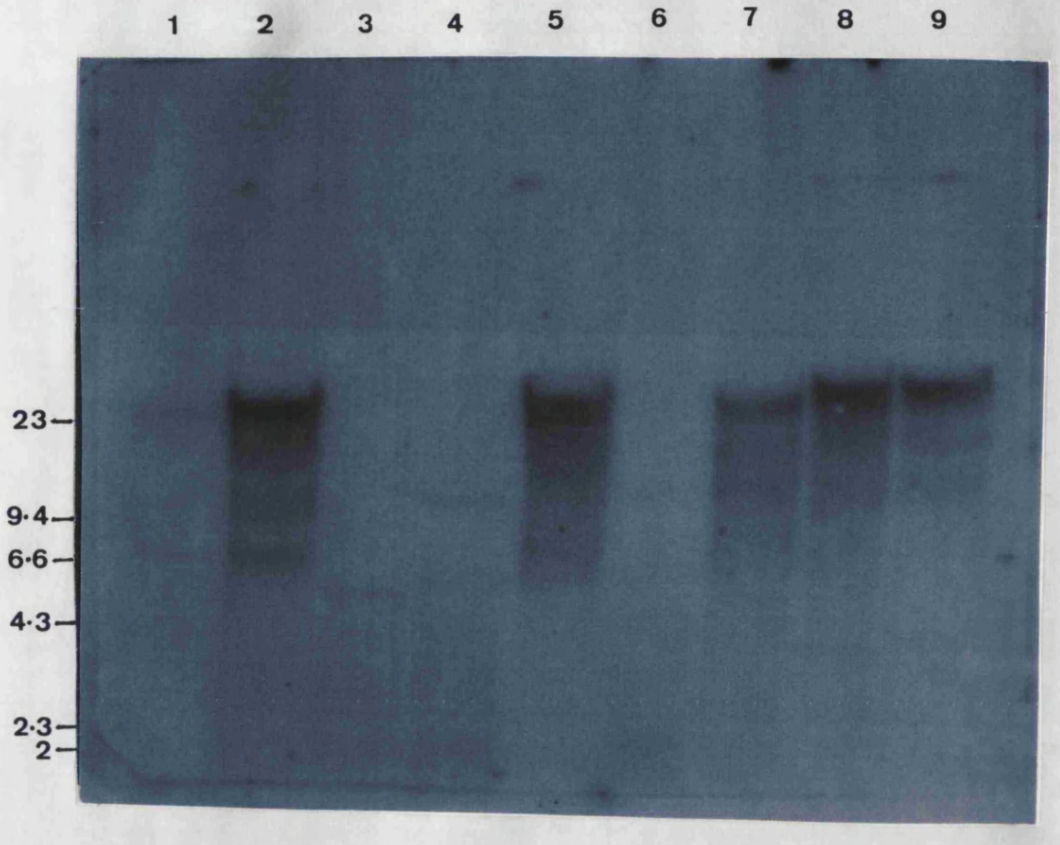

Figure 4.15b Autoradiograph of Southern blot probed with degenerate 17 base oligonucleotide A (30 day exposure)

Lane 1: λ *Hin*dIII
Lane 2: *Hin*dIII / *Sph*I
Lane 3: *Hin*dIII / *Sal*I
Lane 4: *Pst*I / *Sal*I
Lane 5: *Sph*I / *Sal*I
Lane 6: *Sph*I / *Pst*I
Lane 7: *Sma*I / *Eco*RI
Lane 8: *Sma*I / *Bam*HI
Lane 9: *Bam*HI / *Eco*RI

Fragment sizes are indicated in kilobases. Figure 4.15a shows the corresponding agarose gel. The blot was washed at 37°C in 0.5x SSC.

these conditions did not improve the hybridisation specificity (autoradiograph exposed for 30 days).

As a single hybridising restriction fragment could not be identified using oligonucleotide A alone, a further oligonucleotide was designed as described below.

4.8 Screening Southern blots with a non-degenerate 29 base oligonucleotide probe

A 29 base non-degenerate oligonucleotide was designed to the citrate synthase N-terminal sequence, with reference to the codon usage data previously described in Figure 4.1. It's predicted melting temperature is 102°C. The new sequence did not overlap the previous 44 base oligonucleotide and is shown in Figure 4.16.

The non-degenerate design approach was taken for two reasons: the difficulties already encountered when using degenerate oligonucleotides as probes and the lack of suitable regions to design further degenerate oligonucleotides.

The new oligonucleotide was made shorter than the previous 44 base non-degenerate probe to reduce the possibility of hybridisation to sequences other than the citrate synthase gene. The 44 base probe may have been excessively long, contributing to its ability to hybridise to an unrelated gene.

4.8.1 Preliminary dot-blot results

The 29 base probe was observed to hybridise to dot-blots of 1 µg genomic DNA when hybridisation and washing were carried out at 55°C.

```
 G   E   L   K   R   G   L   E   G   V   L V A E S K L S F I D G D A G Q L V Y X G Y D I E D L

GGA GCA CTA AAA CGA GGA CTA GCA GGA GTA
 G   G   G   G   G   G   G   G   G   G
 C   C   C       C   C   C   C   C   C
 T   T   T       T   T   T   T   T   T
       TTA     AGA     TTA
         G       G       G

GGC GAG CTC AAG CGC GGC CTC GAG GGC GT
```

Figure 4.16 Sequence of 29 base oligonucleotide probe

The citrate synthase N-terminal sequence is shown, with all possible codons below each residue. The probe sequence is shown in **bold**.

4.8.2 Southern blot results

Hybridisation was carried out at 55°C to Southern blots of genomic DNA digested with single restriction enzymes. Washing was also carried out at 55°C. Figures 4.17a and 4.17b show the result (autoradiograph exposed for 30 days). The majority of the probe hybridised to high molecular weight restriction fragments or undigested DNA. Smaller hybridising fragments are visible, notably a 4 kb *Eco*RI fragment (Figure 4.17a, lane 7).

As it was necessary to expose each autoradiograph for a long period (30 days) it was considered impractical to use the Southern blotting technique to follow up this result in a reasonable timescale. At these hybridisation and washing stringencies, exposure times were prohibitively long, while use of less stringent conditions would run the risk of introducing non-specific hybridisation without greatly decreasing the necessary exposure time.

From this point onwards the 29 base oligonucleotide was applied as a PCR primer as described in Chapter 6. This technique had the potential to be optimised in a much shorter time.

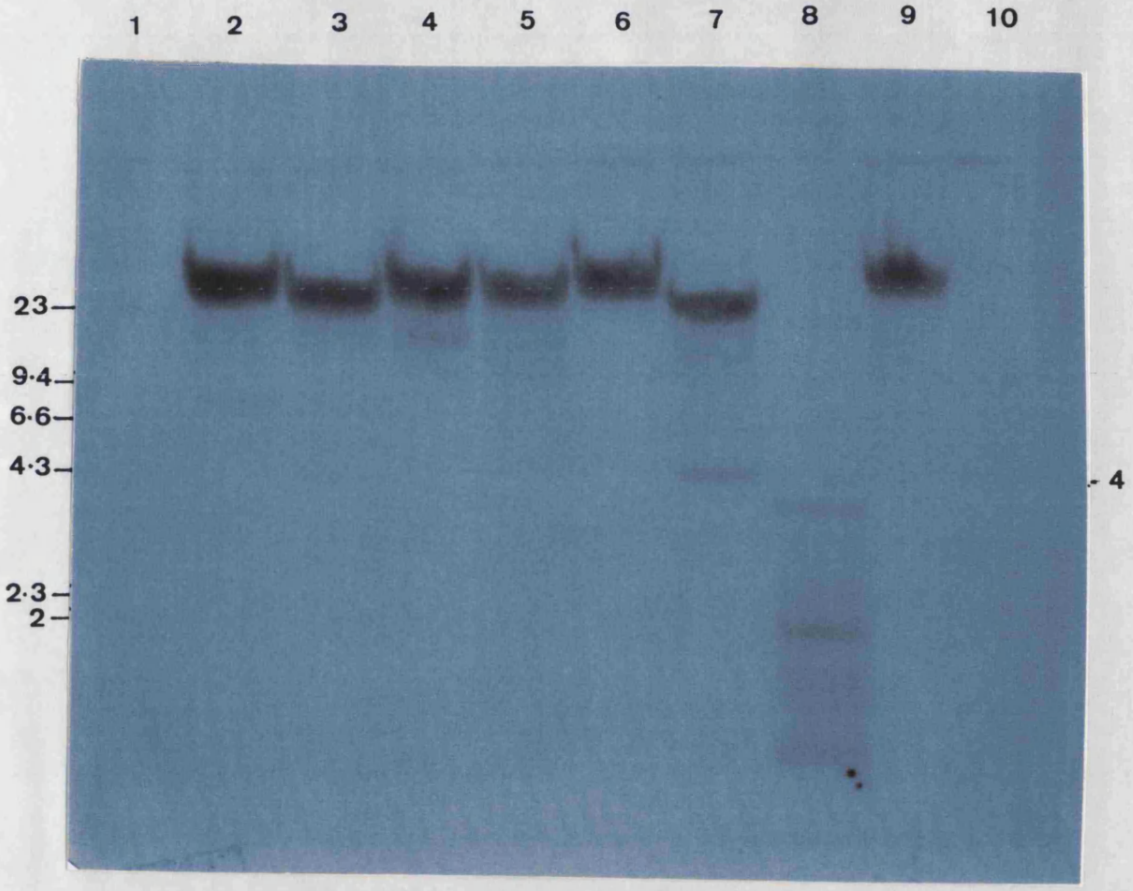

Figure 4.17b Autoradiograph of Southern blot probed with non-degenerate 29 base oligonucleotide (30 day exposure)

Lane 1: λ *Hin*dIII
Lane 2: *Hin*dIII
Lane 3: *Pst*I
Lane 4: *Sph*I
Lane 5: *Sma*I
Lane 6: *Bam*HI
Lane 7: *Eco*RI
Lane 8: *Sal*I
Lane 9: *Kpn*I
Lane 10: λ *Pst*I

Fragment sizes are indicated in kilobases. Figure 4.17a shows the corresponding agarose gel. The blot was washed at 55°C, in 2x SSC.

4.9 Discussion

This chapter describes the hybridisation results obtained with two different types of oligonucleotide probe designed to an N-terminal peptide sequence: long, non-degenerate probes designed with reference to codon usage and shorter, degenerate probes incorporating all possible sequences in the synthetic mixture. The former relies on its length to give specificity, under the assumption that the percentage of mis-matches in its sequence will be low and hybridisation will occur despite their presence. An accurate estimate of codon usage is a pre-requisite. This approach has the advantage that if the probe is a good match, all of the oligonucleotide molecules have the potential to hybridise to the target, as they all have the same sequence.

The latter approach relies on the use of a probe mixture containing all possible sequences, thereby guaranteeing that some are perfect matches to the target. However, only a small percentage of the molecules have the potential to hybridise, reducing the maximum signal that it is possible to obtain.

The 44 base probe hybridises to a non-citrate synthase sequence, a problem that may have been avoided by the use of a shorter oligonucleotide (Figure 4.18). The hybridisation data in Figure 4.5b shows that the binding site of the probe lies within the 1.5 kb *Aval* fragment. When this fragment is cut with *Hin*dIII the strong hybridisation is lost and a small, weakly hybridising *Aval* / *Hin*dIII is released. This fragment is visible in Figure 4.6b (lane 2). From the sequencing data this fragment is 310 base pairs in length.

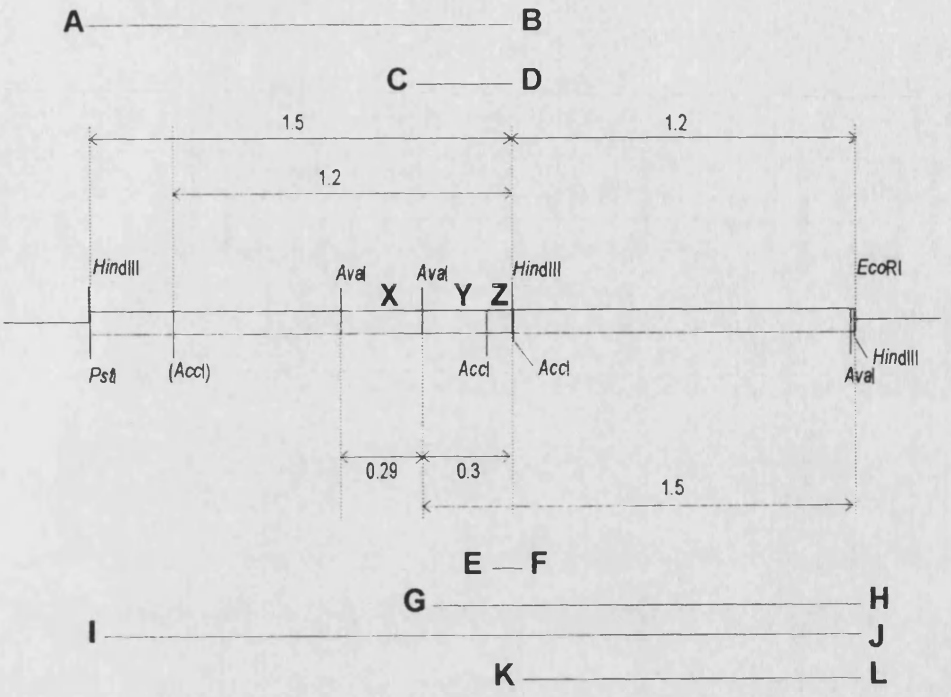

Figure 4.18 Possible probe binding sites within 2.9 kb *Pst*I / *Eco*RI clone

All distances are in kilobases. X, Y and Z indicate the positions of the sites. Their sequences, including identity with the probe, are shown below (the probe would hybridise to the complementary strand):

X (24/44 matches):

870

```
ORF:      ATCGAACTCGCGGAGTCAGCGACCATCGAAGACGGCGCAGTCGT
           ***    * ** **    **  ****** * ** *** * *
Probe:    GTCGCCGACTCGAAGCTTTCGTTCATCGACGGCGACGCCGGCCA
```

Y (25/44 matches):

687

```
ORF:      ATCACGGACGCGATTCAGTCACTCCTCGAAGACGGCTACGCCAT
           ** * *** ***  *  **  ** **** * ** *  ** *
Probe:    GTCGCCGACTCGAAGCTTTCGTTCATCGACGGCGACGCCGGCCA
```

Z (24/44 matches):

507

```
ORF:      GTCGAGAACCCACAGCAGTTCGGCATTGCGGACGTCGACGACCA
          **** * ** *  ***  *    *** *  * ** ** ** ***
Probe:    GTCGCCGACTCGAAGCTTTCGTTCATCGACGGCGACGCCGGCCA
```

*Hin*dIII fragment AB hybridises weakly
*Hin*dIII/*Ava*I fragment CD hybridises weakly
*Acc*I fragment EF fragment hybridises weakly
*Ava*I fragment GH hybridises strongly
*Pst*I/*Eco*RI fragment IJ hybridises strongly
*Hin*dIII fragment KL does not hybridise

subclones were sequenced such that the presence of a citrate synthase gene could be discounted.

The second approach used to locate the citrate synthase gene used mixed pools of probes. It is difficult to predict hybridisation conditions for mixed sequences as their T_m values vary considerably within the pool. Initial experiments were therefore carried out using hybridisation and washing temperatures below the lowest predicted T_m and the stringency of subsequent experiments was gradually increased.

If the perfectly-matched probe has a T_m in the lower part of the pool's T_m range it is unlikely that its specific hybridisation will be retained after all the non-specifically hybridised probes have been washed off. This effect may have contributed to the poor hybridisation results obtained with the pair of 17 base probes. The long exposure times required could suggest inefficient end-labelling of the probe, although approximately 40% or more of the oligonucleotides were labelled in all cases. Normally one would expect about 50% of the oligonucleotides to be labelled. Efforts were made to increase incorporation by re-cleaning the oligonucleotide solutions by ethanol precipitation, increasing the molar excess of ATP over oligonucleotides and using fresh reagents; however no improvement in incorporation was observed.

The second non-degenerate oligonucleotide was designed with reference to the results obtained for the 44 base probe. For the reasons described above, it was made shorter than the previous non-degenerate probe and to a different region of the N-terminal sequence. There was no reason from previous observations to suggest that a second attempt at this approach to probe design would be unsuccessful, provided that these changes were made. Despite reasonable levels of ATP incorporation when

end-labelling (approximately 40% of the oligonucleotides labelled), autoradiograph exposure times were very long. Although the hybridisation appeared to be fairly specific, analysis of further Southern blots and screening of cloned restriction fragments under these conditons would be very time consuming. Subsequent cloning strategies were centred around a PCR approach as this appeared to offer a more efficient way forward for the project.

Chapter 5

Affinity purification of citrate synthase and determination of internal peptide sequence

5.1 Introduction

It is clear from the results presented in Chapter 4 that the identification of a genomic fragment containing the citrate synthase gene by hybridisation to oligonucleotides designed from the N-terminal amino acid sequence met with limited success. Knowledge of a suitable internal peptide sequence would therefore be highly valuable as it provides the opportunity to attempt PCR amplification using primers designed entirely from peptide sequences. An oligonucleotide can be designed from its corresponding peptide sequence with far greater confidence than is possible when designing a consensus probe from a multiple sequence alignment. The chance of successful PCR amplification is greatly enhanced when using such oligonucleotides as primers. The strategies used to design PCR primers are discussed in greater detail in Chapter 6.

Due to lack of material, re-purification of citrate synthase was required before further sequence analysis could be made. The original purification procedure gave a poor recovery of partially pure enzyme. An alternative was sought at this stage in an effort to improve recovery and purity in the final preparation.

Matrex Gel Red A is an agarose support to which the triazine dye ligand Procion Red HE-3B is covalently bound. The structure of Procion

Red is shown in Figure 5.1. This dye has a high affinity for NADP-linked enzymes, but a lower affinity for NAD-linked enzymes. It has found applications in the purification of dehydrogenases where it is thought to bind to the nucleotide-binding site. The stuctural basis of this binding is unknown, although from studies on lactate dehydrogenase (Kirchberger *et al.*, 1989) it appears that the rigid conformation of the terminal rings and the presence of the terminal ring sulphonate groups are important. Coenzyme A shares structural similarities with NADP, both having an adenine moiety linked to a phosphorylated ribose, although the two are phosphorylated at different positions on the sugar.

Based on this information, studies have previously been carried out on the binding of the coenzyme A-linked enzymes citrate synthase and succinate thiokinase, from six different organisms (Weitzman & Ridley, 1983). Significant purifications were achieved for both enzymes using a simple salt gradient elution.

Matrex Gel Red A has been used with success in the purification of citrate synthase from *Acinetobacter calcoaceticus* (Mitchell & Weitzman, 1983) and *Bacillus coagulans* (Schendel *et al.,* 1992), eluting ligand-bound enzyme with either a salt gradient or biospecifically with oxaloacetate and coenzyme A. More recently the citrate synthases of a number of Archaea, including *Sulfolubus solfataricus* (James *et al.*, 1994) and *Pyrococcus furiosus* (Muir *et al.* 1994a) have also been purified using this ligand. This method therefore appeared to be a good candidate for purification of the *Hf.volcanii* citrate synthase. Salt gradients (0-800 mM KCl in the case of *B.coagulans* above) have been found to elute bound citrate synthase from Red A; however, given that halophilic citrate synthases only achieve native

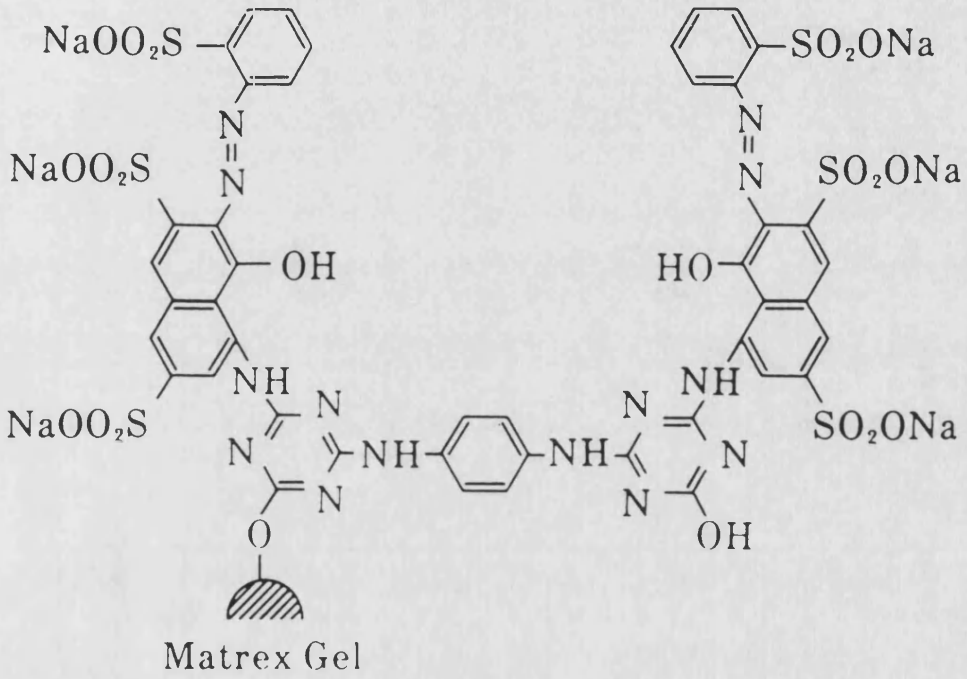

Figure 5.1 Partial structure of Matrex Gel Red A

conformation in high salt conditions the possibility of specific binding and elution in high salt warranted investigation.

6.2 Outline of the purification scheme

Hf.volcanii (2 g wet weight) was resuspended in 10 ml 20 mM Tris-HCl (pH 8.0), 2 mM EDTA, 0.1 M KCl and a cell extract prepared by sonication as described in Section 2.1.1.2.

The cell extract was applied to a 10 ml column of Matrex Gel Red A equilibrated with the same buffer. After washing, citrate synthase was eluted with 1 mM oxaloacetate, 0.2 mM coenzyme A in the same buffer. Fractions containing citrate synthase activity were pooled and concentrated to a total volume of 2 ml using solid polyethylene glycol as described in Section 2.2.3. This method was found to give > 90% recovery of citrate synthase activity after concentration.

Microsep microconcentrators (10 kD cutoff) were also tried as a method of protein concentration but proved unsuitable as the flow rate through the filter decreased rapidly after the first 10 minutes of centrifugation, suggesting that the protein was adhering to the filter membrane. Only 50% of the citrate synthase activity applied to the concentrator could been recovered when using this method.

The concentrated solution was adjusted to 2 M KCl by addition of solid KCl and was then applied to an FPLC Superdex-200 gel filtration column equilibrated with 20 mM Tris-HCl (pH 8.0), 2 mM EDTA, 2 M KCl. Fractions containing citrate synthase activity were pooled, dialysed against 10 mM ammonium bicarbonate, 0.01% (w/v) SDS and dried in a Savant Speedvac concentrator.

5.3 Cyanogen bromide cleavage

Cyanogen bromide cleavage and N-terminal sequencing were carried out by Mrs. Janice Young at Zeneca Pharmaceuticals, Macclesfield. An estimated 20 µg protein was dissolved in 250 µl 70% (v/v) formic acid, in a polypropylene tube and 186 µl cyanogen bromide solution (10 mg ml^{-1} in 70% (v/v) formic acid) added. After purging with nitrogen, the sample was incubated in the dark for 24 hours, at room temperature.

The sample was dried, after addition of 5 µl 10% (w/v) SDS to prevent fragments sticking to the tube. The sample was re-dissolved in water and dried again to remove the acid. The sample was dissolved in reduced tris-tricine sample buffer, run on a 16.5% tris-tricine polyacrylamide minigel and blotted onto Problott membrane. A band of approximately 6 kD was cut out.

5.4 N-terminal sequencing

To repeat the N-terminal sequencing of the intact protein, 5-10 µg was dissolved in reducing sample buffer, subjected to SDS-polyacrylamide gel electrophoresis and blotted onto Problott membrane. The single band was cut out and sequenced using an Applied Biosystems 477 Sequencer (modified to run blot cycles) with a gas phase TFA delivery.

The cyanogen bromide fragment was sequenced using an Applied Biosystems 475 Sequencer. In both cases the sequencers were coupled to an Applied Biosystems 120A Phenylthiohydantoin analyser.

5.5 Results

5.5.1 Purification

The purification is summarised in the table below

Step	Protein mg	CS Activity U	Sp Activity U mg^{-1}	Purification	Yield
Cell-free extract	160	13.5	0.085	1	100
Red A	0.66	10.8	16.42	196	81
PEG concentration	-	10.7	-	-	-
Superdex-200	0.21	9.5	45.2	532	70

Figures 5.2 and 5.3 show elution profiles of the Red A column and Superdex-200 column respectively. Figure 5.4 shows an SDS-polyacrylamide gel of the cell extract, the pooled fractions from the Red A column and the pooled fractions from the Superdex-200 column after dialysis and drying.

CoA in the Red A gel fractions was assayed by addition of DTNB to an aliquot of the fraction, in the absence of oxaloacetate. The reaction was followed to completion spectrophotmetrically as for a citrate synthase assay and the CoA concentration calculated from the final absorbance at 512 nm.

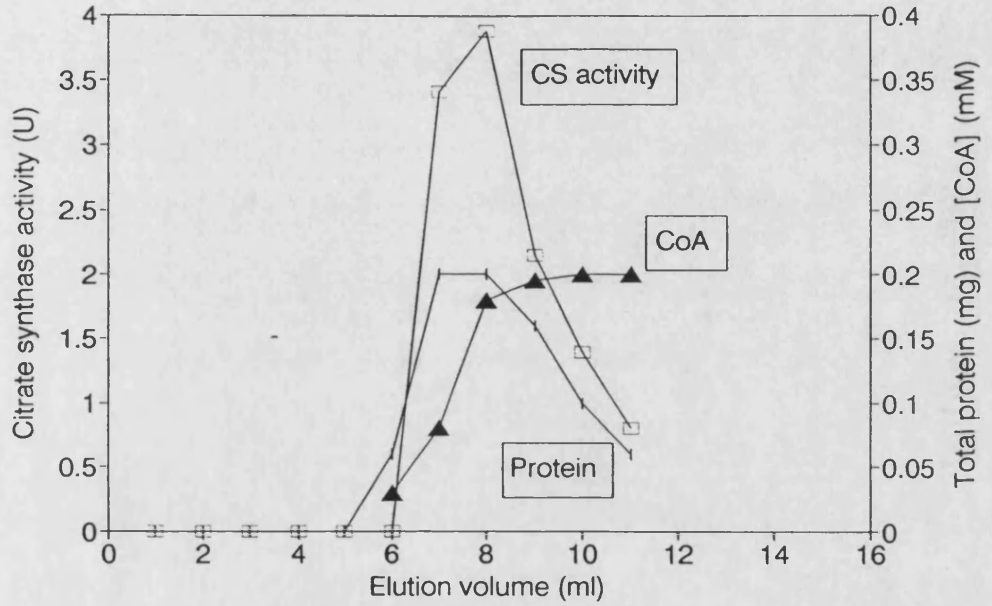

Figure 5.2 Matrex gel Red A chromatography of citrate synthase

The elution profile of a 10 ml column is shown. Citrate synthase activity (U), total protein (mg) and Coenzyme A concentration (mM) in each 1 ml fraction are plotted against elution volume (ml).

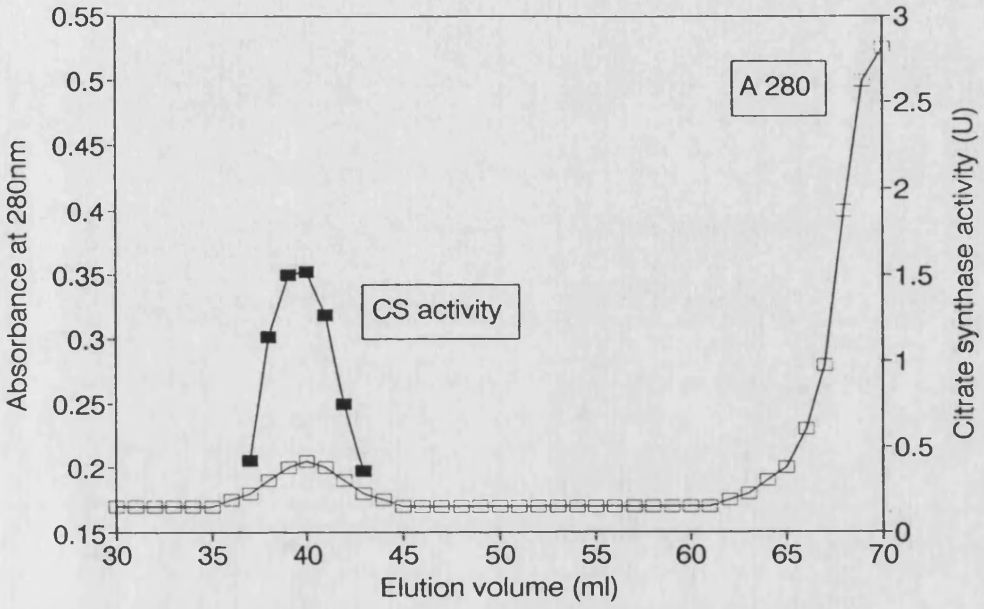

Figure 5.3 FPLC Superdex-200 gel filtration of citrate synthase

Citrate synthase activity (U) and absorbance at 280 nm are plotted against elution volume (ml) for each 1 ml fraction. The void volume of the column is 30 ml. The absorbance peak at 60 - 70 ml is caused by CoA eluting from the column.

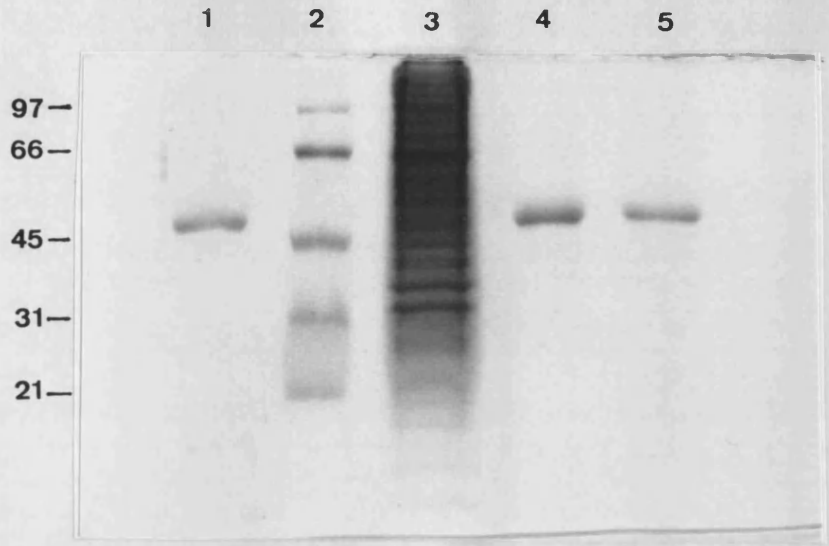

Figure 5.4 SDS-PAGE of citrate synthase

Lane 1: Pig heart citrate synthase
Lane 2: Molecular weight markers
Lane 3: *Hf.volcanii* cell extract
Lane 4: Matrex gel Red A eluent
Lane 5: Superdex-200 eluent

Sizes are indicated in kD.

5.5.2 N-terminal sequencing

The re-sequencing of the N-terminus of the uncleaved protein confirmed the 38 residue sequence previously obtained. The yield of fragments from cyanogen bromide treatment was unexpectedly low: on SDS-PAGE only a single band 6 kD band was visible. The N-terminal sequence of the 6 kD band was determined and is shown below. Residues in parentheses are tentative assignments and X indicates an unknown residue.

6 kD XIPIDLYTPIFAVSRAG(G)

5.6 Discussion

In terms of final purity and recovery of citrate synthase the revised purification scheme is a significant improvement on that previously used. The dried citrate synthase appeared pure as judged by the appearance of a single band on SDS gels. The most significant loss of material occurred during the concentration step, where the protein appeared to bind to the filtration membrane.

Trials of the Red A column were carried out on small (1 ml bed volume) columns using a variety of different loading and elution conditions. Initially the procedure described above was carried out in buffers containing 1 M KCl, the concentration at which the enzyme has maximum activity, since if the Red A dye is acting as nucleotide analogue it is likely

to bind at the acetyl coenzyme A binding site and a fully active enzyme ensures correct conformation of that site. Under these conditions citrate synthase bound irreversibly to the column, resisting elution with concentrations of oxaloacetate and coenzyme A as high as 10 mM and 1 mM respectively. The binding appeared to be specific as the vast majority of the protein loaded onto the column could be accounted for in the initial wash.

Reducing the KCl concentration to 0.1 M, the minimum KCl concentration at which the enzyme retains both activity and stability, resulted in modified specific binding which could be overcome by addition of 1 mM oxaloacetate and 0.2 mM coenzyme A as described. The difference in binding could be due to non-specific interactions between citrate synthase and the dye-ligand matrix which occur in addition to the specific binding, in the presence of 1 M KCl. Alternatively the change in binding with KCl concentration may be caused by a change in the enzyme's affinity for the ligand, coenzyme A or oxaloacetate.

It has since been shown in our laboratory that as the KCl concentration is decreased from 1 M to 0.1 M, the K_m of oxaloacetate decreases and the K_m of acetyl coenzyme A increases (Henderson, 1994). This suggests that oxaloacetate may become a more efficient eluant at lower salt concentrations, where its K_m is lower. If the K_m of coenzyme A increases at lower KCl concentrations, the binding to the column may be weaker under these conditions.

The internal peptide sequence was aligned by hand to the same citrate synthase multiple sequence alignment as was used to align the N-ternimal sequence. Figure 5.5 shows the relevant region of the alignment. The sequence shows very high identity with the other citrate synthases

and includes four residues that are conserved in all the enzymes. The cyanogen bromide cleavage site is likely to be a methionine in the equivalent position to that observed in the *E.coli*, *A.aceti*, *R.prowazekii*, *M.smegmatis* and *Tp.acidophilum* sequences. These sequence data were used to design an oligonucleotide primer for use in PCR as described in Chapter 6.

```
                    401                                                 450
     Chicken heart  aanpwPNvDa hsgvllqyyg mt.emnyYTv lFgvsRalGv laqli.wsra
        Pig kidney  aknpwPNvDa hsgvllqyyg mt.emnyYTv lFgvsRalGv laqli.wsra
Yeast (cytoplasm)  tknpwPNvDa hsgvllqyyg lk.essfYTv lFgvsRafGi laqli.tdra
   Yeast (mitoch)  tknpwPNvDs hsgvllqyyg lt.easfYTv lFgvaRaiGv lpqli.idra
      Tetrahymena  ianpyPNvDc hsgvllyslg lt.eyqyYTv vFavsRalGc manli.wsra
     Mycobacterium ..rikPNlD. fptgayylmd fpi..esFTp lFvmsRitGw tahimeqaas
         Bacillus  ..glfPNlDy yaapvywklg ipi..plYTp iFfssRtvGl cahvmeqhen
     Acinetobacter ..klyPNvDf ysgiilkaig ip..temFTv iFalaRtvGw ishwlemhsg
      Pseudomonas  ..nlyPNvDf ysgiilkaig ip..tsmFTv iFalaRtvGw ishwqemlsg
      Escherichia  ..klyPNvDf ysgiilkamg ip..ssmFTv iFamaRtvGw iahwsemhsd
      Acetobacter  ..klyPNvDf ysgiilkamg ip..tsmFTv lFavaRttGw vsqwkemiee
         Coxiella  ..klyPNvDf ysgltlnaig ip..snmFTv iFalsRtvGw ishwmemmss
       Rickettsia  ..klyPNvDf ysgiiykamg ip..sqmFTv lFaiaRtvGw maqwkemhed
     Thermoplasma  ..giyPNtDy fsgivymsig fplrnniYTa lFalsRvtGw qahfieyvee
        Haloferax                        IP  IDLYTP IFAVSRAG
        Consensus  -----PN-D- ---------- --------T- -F---R--G- ----------
```

Figure 5.5 Alignment of 6kD peptide N-terminal sequence with citrate synthase multiple sequence alignment

Chapter 6

Amplification of the citrate synthase gene of *Haloferax volcanii* from genomic DNA by PCR

6.1 Introduction

The amplification of genomic DNA by the polymerase chain reaction provides an alternative to the Southern blot screening approach in gene cloning. The success of a PCR reaction relies heavily on the specificity of the primers used, which in turn requires careful primer design at the planning stage.

Primers may be designed with reference to a known amino acid sequence and codon-usage table, or alternatively an alignment of related protein sequences may be created and used to identify highly conserved regions of sequence, which are in turn used for primer design. The first method is far superior as it relies on only one set of assumptions (the codon-usage table), while the second relies on two (both the primary amino acid sequence and the codon use are inferred). Both of these strategies were applied in this project.

6.2 PCR amplification of an 830 bp sequence from genomic DNA using N-terminal and consensus amino acid sequences for primer design

The aim of the reactions described in this section was to amplify a large part of the citrate synthase gene for cloning and sequencing. From these sequence data a restriction map could be made and long, perfectly matched

oligonucleotide probes designed with which to screen Southern blots as described in Chapter Four, but with much greater efficacy.

6.2.1 Primer design

These reactions were carried out before the internal protein sequence data described in Chapter Five had been determined. The sense primer used was the non-degenerate 29 base oligonucleotide made to the citrate synthase N-terminal amino acid sequence (described in Figure 4.16).

The anti-sense primer was therefore made with reference to an alignment of citrate synthase amino acid sequences, shown in Figure 6.1. The alignment was created by M. McCormack, in our laboratory using the program PILEUP in the GCG program package. There are four particularly well conserved regions apparent in the alignment (designated R1 to R4 in Figure 6.1), each containing a number of residues conserved in all citrate synthases so far sequenced.

The only Archaeal citrate synthase sequence available is that of *Thermoplasma acidophilum* (Sutherland *et al.*, 1991) which shows 19-20% identity with eukaryotic sequences and 27-28% with bacterial sequences. The citrate synthase sequences currently available show higher identity within a domain than between domains (Sutherland *et al.*, 1991). The N-terminal sequence of the *Hf.volcanii* enzyme shows highest identity with the corresponding region of the *E.coli* enzyme (50%) and the *Tp.acidophilum* enzyme (42%). This was considered in choosing residues from the conserved regions of the alignment (R1 to R4) most likely to occur in the *Hf.volcanii* enzyme, as shown in Figure 6.2.

The most probable DNA sequences corresponding to these regions were estimated using the codon use data in Figure 4.1 and are given in Figure 6.3.

```
                    1                                                    50
    Chicken heart   ...asstnlk dvlaalipke qariktfrqq hggtalgqit v.dms.....
       Pig kidney   ...asstnlk diladlipke qariktfrqq hgntvvgqit v.dmm.....
Yeast (cytoplasm)   ...sqektlk erfseiypih aqdvrqfvke hgktkisdvl l.eqv.....
   Yeast (mitoch)   ...aseqtlk erfaeiipak aqeikkfkke hgktvigevl leeqa.....
      Tetrahymena   ....sqtnlk kviaeiipqk qaelkevkek ygdkvvgqyt v.kqv.....
     Mycobacterium  .......apq qalrtaalrr lqpvrlrhlq lrraggesmt tateseapri
         Bacillus   .......... .......... .......... .......... ....vntnqf
     Acinetobacter  ....seatgk kavlhldgke .ielpiysgt lgpdvidvkd vlas.ghftf
       Pseudomonas  .......adk kaqliiegsa pvelpvlsgt mgpdvvdvrg ltat.ghftf
       Escherichia  .......adt kakltlngdt aveldvlkgt lgqdvidirt lgsk.gvftf
       Acetobacter  sasqkegkls tatisvdgks a.empvlsgt lgpdvidirk lpaqlgvftf
          Coxiella  .......snr kaklsfenqs .vefpiyspt lgkdvidvkt l.gnhgayal
        Rickettsia  .tngnnnnle faelkirgk. lfklpilkas igkdvidisr vsaeadyfty
     Thermoplasma   .......... .......... .......... .......... ....peteei
          Consensus ---------- ---------- ---------- ---------- ----------

                    51                                                   100
    Chicken heart   ygGmrgmkgl vyetsvldpd egi.rfrGfs ipecqkllpk gg......xg
       Pig kidney   ygGmrgmkgl vyetsvldpd egi.rfrGys ipecqkmlpk ak......gg
Yeast (cytoplasm)   ygGmrgipgs vwegsvldpe dgi.rfrGrt iadiqkdlpk ak......gs
   Yeast (mitoch)   ygGmrgikgl vwegsvldpe egi.rfrGrt ipeiqrelpk ae......gs
      Tetrahymena   igGmrgmkgl msdlsrcdpy qgi.ifrGyt ipqlkeflpk adpkaadqan
     Mycobacterium  hkGlagvvvd ttaiskvvpe tnsltyrGyp vqd.l..... .........a
         Bacillus   ipGlegvias etkisfldtv nseivikGyd lla.l..... .........s
     Acinetobacter  dpGfmatasc eskitfidgd kgillhrGyp idq.l..... .........a
       Pseudomonas  dpGfmstasc eskityidgd kgvllhrGyp ieq.l..... .........a
       Escherichia  dpGftstasc eskitfidgd egillhrGfp idq.l..... .........a
       Acetobacter  dpGygetaac nskitfidgd kgvllhrGyp iaq.l..... .........d
          Coxiella  dvGfystaac eskitfidge kgillyrGyp idq.l..... .........a
        Rickettsia  dpGfmstasc qstityidgd kgilwyrGyd ikd.l..... .........a
     Thermoplasma   skGledvnik wtrlttidgn kgilrygGys vedii..... .........a
          Consensus --G------- -------d-- -g----rG-- ---------- ----------

                    101                                                  150
    Chicken heart   geplpeglfw lLvtGqiPtg aqvswLskew akraalpshv vtmldnFptn
       Pig kidney   eeplpeglfw lLvtGqiPte eqvswLskew akraalpshv vtmldnFptn
Yeast (cytoplasm)   sqplpealfw lLltGevPtq aqvenLsadl msrselpshv vqlldnLpkd
   Yeast (mitoch)   teplpealfw lLltGeiPtd aqvkaLsadl aarseipehv iqlldsLpkd
      Tetrahymena   qeplpegifw lLmtGqlPth aqvdaLkhew qnrgtvnqdc vnfilnLpkd
     Mycobacterium  .aqcsfeqvy lLwhGelPtd .qlalFsqre rasrridrsm qallakLpdn
         Bacillus   ktkgyldivh lLleGtiPne aekqhLeetl kqeydvpdei iqvlslLpkt
     Acinetobacter  tqadyletcy lLlnGelPta eqkveFdakv rahtmvhdqv srffngFrrd
       Pseudomonas  eksdyletcy lLlnGelPta aqkeqFvgti knhtmvheql ktffngFrrd
       Escherichia  tdsnylevcy iLlnGekPtq eqydeFkttv trhtmiheqi trlfhaFrrd
       Acetobacter  enasyeeviy lLlnGelPnk vqydtFtntl tnhtllheqi rnffngFrrd
          Coxiella  dksdymevcy lLmyGelPnk gekekFvrti kehtsvyeqv tkffngFhyd
        Rickettsia  eksdflevay lMiyGelPss dqycnFtkkv ahhslvnerl hylfqtFcss
     Thermoplasma   sgaqdeeiqy lFlyGnlPte qelrkYketv qkgykipdfv inairqLpre
          Consensus ---------- lL--G--P-- ---------- ---------- ----------
```

Figure 6.1 Multiple sequence alignment of citrate synthases.

```
                    151                                                 200
    Chicken heart   lhpmsqlsaa itaLnsesnf arayaeg.il rtkywemvye samdLiaklp
       Pig kidney   lhpmsqlsaa itaLnsesnf arayaeg.ih rtkyweliye dcmdLiaklp
Yeast (cytoplasm)   lhpmaqfsia vtaLeseskf akayaqg.is kqdywsytfe dsldLlgklp
   Yeast (mitoch)   lhpmaqfsia vtaLeseskf akayaqg.vs kkeywsytfe dsldLlgklp
      Tetrahymena   lhsmtmlsma llyLqkdskf aklydegkis kkdywepfye dsmdLiakip
     Mycobacterium  chpmdvvrta isyLg..... aedleedvdt aeanya.... kslrMfavlp
         Bacillus   ahpmdalrtg vsvLa..... sfdtellnre hstnlk.... rayqLlgkip
     Acinetobacter  ahpmaimvgv vgaLs..... afyhnnldie dinhrei... tairLiakip
       Pseudomonas  ahpmavmcgv igaLs..... afyhdsldit npkhrqv... sahrLiakmp
      Escherichia   shpmavmcgi tgaLa..... afyhdsldvn nprhrei... aafrLlskmp
      Acetobacter   ahpmailcgt vgaLs..... afypdandia ipanrdl... aamrLiakip
         Coxiella   ahpmamvlst igaLs..... afyhdaldit kpadrel... sairLiakmp
       Rickettsia   shpmaimlaa vgsLs..... afypdllnfn etdy.el... tairMiakip
     Thermoplasma   sdavamqmaa vaaMa..... as.etkfkwn kdtdrdv... aa.eMigrms
         Consensus  -h-m------ ---L------ a--------- ---------- -------k-p

                    201                                                 250
    Chicken heart   cvaakiyrnl yragssigai dskldWshnF tnmlgyt... .....daqft
       Pig kidney   cvaakiyrnl yregssigai dskldWshnF tnmlgyt... .....daqft
Yeast (cytoplasm)   viaakiyrnv fkdg.kmgev dpnadYaknL vnligsk... .....dedfv
   Yeast (mitoch)   viaskiyrnv fkdg.kitst dpnadYgknL aqllgye... .....nkdfi
      Tetrahymena   rvaaiiyrhk yrdsklids. dskldWagnY ahmmgfe... .....qhvvk
     Mycobacterium  tivatdir.. rrqgltpipp hsqlgYaqnF lnmcfgevpe ......pvvv
         Bacillus   nivansyh.. ilhseepvqp lqdlsYsanF lymitgkkpt ......elee
     Acinetobacter  tlaawsyk.. ytvgqpfiyp rndlnYaenF lhmmfatpad rdykvnpvla
       Pseudomonas  tiaamvyk.. yskgepmmyp rndlnYaenF lhmmfntpce tk.pispvla
      Escherichia   imaamcyk.. ysigqpfvyp rndlsYagnF lnmmfstpce p.yevnpile
      Acetobacter   tiaawayk.. ytqgeafiyp rndlnYaenF lsmmfarmse p.ykvnpvla
         Coxiella   tlaamsyk.. ysigqpfmhp rramnYaenF lhmlfgtpye et.epdpvla
       Rickettsia   tiaamsyk.. ysigqpfiyp dnsldFtenF lhmmfatpc. tkykvnpiik
     Thermoplasma   aitvnvyr.. himnmpaelp kpsdsYaesF lnaafgrkat ke......ei
         Consensus  ------y--- ---------- ---------- ---------- ----------
                                                                  ▼
                                                                  ●●
                    251                                         300
    Chicken heart   elmrlyLtih sDHeggnvSa htshlvgSal sdpYlsfaaa mngLaGplHG
       Pig kidney   elmrlyLtih sDHeggnvSa htshlvgSal sdpYlsfaaa mngLaGplHG
Yeast (cytoplasm)   dlmrlyLtih sDHeggnvSa htshlvgSal sspYlslasg lngLaGplHG
   Yeast (mitoch)   dlmrlyLtih sDHeggnvSa htthlvgSal sspYlslaag lngLaGplHG
      Tetrahymena   ecirgyLsih cDHeggnvSa htthlvgSal sdpYlsysag vngLaGplHG
     Mycobacterium  rafeqsMvly aEHs.fnaSt faarvvtStq sdiYsavtaa igaLkGslHG
         Bacillus   kifdrsLvly sEHe.lpnSt ftarviaStl sdlYgaltga vasLkGhlHG
     Acinetobacter  ramdriFtlh aDHe.qnaSt stvrlagStg anpYacisag isaLwGpaHG
       Pseudomonas  kamdriFilh aDHe.qnaSt stvrlagSsg anpFaciasg iaaLwGpaHG
      Escherichia   ramdriLilh aDHe.qnaSt stvrtagSsg anpFaciaag iasLwGpaHG
      Acetobacter   ramnriLilh aDHe.qnaSt stvrlagStg anpFaciaag iaaLwGpaHG
         Coxiella   ramdriFilh aDHe.qnaSt ttvrvagStg anpFacisag isaLwGpaHG
       Rickettsia   nalnkiFilh aDHe.qnaSt stvriagSsg anpFacistg iasLwGpaHG
     Thermoplasma   damntaLily tDHe.vpaSt taglvavStl sdmYsgitaa laaLkGplHG
         Consensus  ---------- -EHe--n-S- -------S-- ---------- ---L-Gp-HG
                                                                 ←----R1-
```

Figure 6.1 Multiple sequence alignment of citrate synthases.

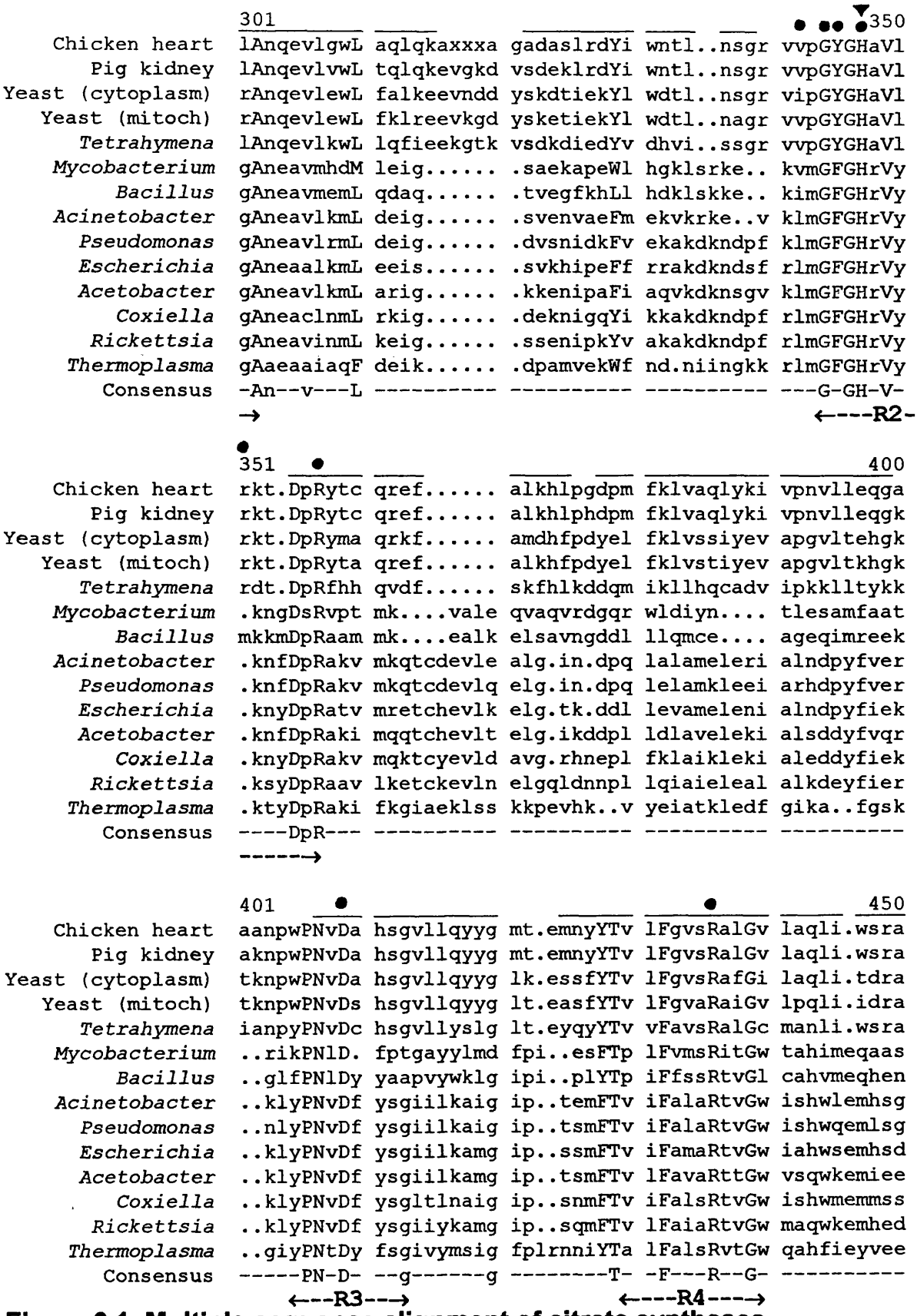

Figure 6.1 Multiple sequence alignment of citrate synthases.

```
                 451   •       ___    ___  __     478
    Chicken heart  lgfpleRPks mstd...gli al......
       Pig kidney  lgfpleRPks mstd...gli klvdsk..
Yeast (cytoplasm)  igasieRPks ystekykelv knieskl.
   Yeast (mitoch)  vgapieRPks fstekykelv kkieskn.
      Tetrahymena  fglpieRPgs adlkwfhdky re......
     Mycobacterium ..naliRPls eysgqpqrsl v.......
         Bacillus  ..nrifRPrv lytgarnlrv ed......
     Acinetobacter p.ykigRPrq lytgevqrdi kr......
       Pseudomonas p.ykigRPrq lytghtqrdf talkdrg.
       Escherichia g.mkiaRPrq lyigyekrdf ksdikr..
       Acetobacter pgqrisRPrq lyigapqrdy vplakr..
          Coxiella pdhrlaRPrq lytgeterev isldkrqa
        Rickettsia peqkisRPrq lytgyvhrey kciverk.
      Thermoplasma q.qrliRPra vyvgpaerky vpiaerk.
         Consensus ------RP-- ---------- --------
```

Figure 6.1 Multiple sequence alignment of citrate synthases.

Residues with Dayhoff matches of 1.2 or greater are in Upper Case (Schwartz & Dayhoff, 1979). The Dayhoff match between two amino acids reflects the probability of changing from one amino acid to any other, taking into account factors such as the required change in codon. It's value may vary from -1.2 to 1.5, where 1.5 indicates identity.

The consensus contains residues that occur in at least 12 of the 14 sequences, regardless of their relative Dayhoff values.

Regions of secondary structure in the Pig enzyme are indicated by the horizontal bar above the alignment. Active site residues (▾) and substrate-binding residues (•) in the Pig enzyme are indicated.

Conserved regions considered for PCR primer design are marked R1 to R4.

Region 1:

```
                      300
     Chicken heart LaGplHG lA
       Pig kidney  LaGplHG lA
Yeast (cytoplasm)  LaGplHG rA
   Yeast (mitoch)  LaGplHG rA
      Tetrahymena  LaGplHG lA
     Mycobacterium LkGslHG gA
          Bacillus LkGhlHG gA
     Acinetobacter LwGpaHG gA
       Pseudomonas LwGpaHG gA
       Escherichia LwGpaHG gA
       Acetobacter LwGpaHG gA
          Coxiella LwGpaHG gA
         Consensus L-G--HG gA

     (Haloferax?) LwGpaHG gA
```

Region 2:

```
                        350
     Chicken heart  GYGHaVl rkt.DpR
       Pig kidney   GYGHaVl rkt.DpR
Yeast (cytoplasm)   GYGHaVl rkt.DpR
   Yeast (mitoch)   GYGHaVl rkt.DpR
      Tetrahymena   GYGHaVl rdt.DpR
     Mycobacterium  GFGHrVy .kngDsR
          Bacillus  GFGHrVy mkkmDpR
     Acinetobacter  GFGHrVy .knfDpR
       Pseudomonas  GFGHrVy .knfDpR
       Escherichia  GFGHrVy .knyDpR
       Acetobacter  GFGHrVy .knfDpR
        Rickettsia  GFGHrVy .ksyDpR
     Thermoplasma   GFGHrVy .ktyDpR
         Consensus  G-GH-V------D-R
     (Haloferax?)   GFGHrVy .k?yDpR
```

Region 3:

```
                401
     Chicken heart  wPNvDa hsg
       Pig kidney   wPNvDa hsg
Yeast (cytoplasm)   wPNvDa hsg
   Yeast (mitoch)   wPNvDs hsg
      Tetrahymena   yPNvDc hsg
     Mycobacterium  kPNlD. fpt
          Bacillus  fPNlDy yaa
     Acinetobacter  yPNvDf ysg
       Pseudomonas  yPNvDf ysg
       Escherichia  yPNvDf ysg
       Acetobacter  yPNvDf ysg
          Coxiella  yPNvDf ysg
        Rickettsia  yPNvDf ysg
     Thermoplasma   yPNtDy fsg
         Consensus  -PN-D- ---

     (Haloferax?)   yPNvDf ysg
```

Region 4:

```
                        431
     Chicken heart  Tv lFgvsRalG
       Pig kidney   Tv lFgvsRalG
Yeast (cytoplasm)   Tv lFgvsRafG
   Yeast (mitoch)   Tv lFgvaRaiG
      Tetrahymena   Tv vFavsRalG
     Mycobacterium  Tp lFvmsRitG
          Bacillus  Tp iFfssRtvG
     Acinetobacter  Tv iFalaRtvG
       Pseudomonas  Tv iFalaRtvG
       Escherichia  Tv iFamaRtvG
       Acetobacter  Tv lFavaRttG
          Coxiella  Tv iFalsRtvG
        Rickettsia  Tv lFaiaRtvG
     Thermoplasma   Ta lFalsRvtG
         Consensus  T- -F---R--G

     (Haloferax?)   Tv iFa??Rt?G
```

Figure 6.2 Regions conserved in all known citrate synthase sequences

The possible corresponding sequences in *Hf.volcanii* are indicated. Question marks indicate positions where it was not possible to determine a likely residue for *Hf.volcanii*. Note that only completely conserved residues have been retained in the consensus for this analysis.

Region 1:

	L	w	G	p	a	H	G	q	A
5'-	CTC	TGG	GGC	CCC	GCC	CAC	GGC	GGC	GC

88% G+C

Region 2:

	G	F	G	H	r	V	y	k
5'-	GGC	TTC	GGC	CAC	CGC	GTC	TAC	AA

65% G+C

Region 3:

	y	P	N	v	D	f	y	s	q	
5'-	TAC	CCC	AAC	GTC	GAC	TTC	TAC	TCG	GG	
Antisense	ATG	GGG	TTG	CAG	CTG	AAG	ATG	AGC	CC	-5'

52% G+C

Figure 6.3 Predicted DNA sequences corresponding to conserved regions in citrate synthase amino acid sequences

Region 3 alone was chosen for PCR primer design for the reason described in Section 6.2.1. The sequence of the anti-sense primer synthesised in shown above.

Region 1 was eliminated as the predicted DNA sequence was greater than 85% G+C as such a high G+C content is likely to cause secondary structure formation, primer-dimer artefacts during PCR and possibly non-specific binding of the primer. Region 4 was also eliminated because some positions in the predicted amino acid sequence were highly uncertain. Region 3 was chosen over region 2 because a slightly longer primer could be made and a larger fragment of the gene could be amplified. A 26 base anti-sense consensus primer of the sequence given in Figure 6.3 was synthesised.

The *Tp.acidophilum* citrate synthase gene is 1152 bp in length, therefore the expected length of the PCR product is approximately 200 bp less; about 960 bp.

6.2.2 PCR amplification with N-terminal and consensus primers

PCR reactions were carried out using 80 ng genomic DNA as template, 0.4 µM primers and 50 µM dNTPs. Thirty cycles were performed, melting at 94°C (1 minute), annealing at 55°C (1 minute 30 seconds) and extending at 72°C (2 minutes). The magnesium concentration in the reactions was varied from 0.5 to 5 mM. Figure 6.4 shows examples of the PCR products, separated by electrophoresis on a 1% agarose gel.

A PCR fragment of 830 bp is apparent in the 1 to 5 mM Mg^{2+} lanes. This is approximately 130 bp shorter than expected (corresponding to a citrate synthase some 40 residues shorter than that of *Tp.acidophilum*), and was unlikely to be an amplified citrate synthase sequence, even though control reactions carried out in 1.5 mM Mg^{2+} using single primers confirmed that this band was not an artefact caused by one primer binding at multiple sites (Figure 6.5).

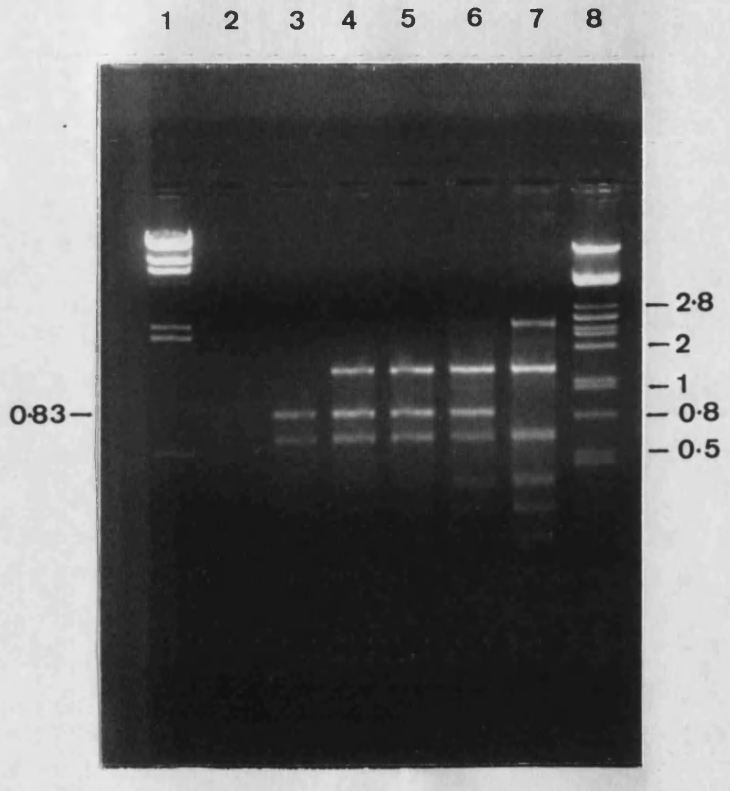

Figure 6.4 PCR amplification of genomic DNA using 29 base N-terminal and 26 base consesus primers, annealing at 55°C

Lane 1: λ *Hin*dIII
Lane 2: 0.5 mM Mg^{2+}
Lane 3: 1.0 mM Mg^{2+}
Lane 4: 1.5 mM Mg^{2+}
Lane 5: 2.0 mM Mg^{2+}
Lane 6: 3.0 mM Mg^{2+}
Lane 7: 5.0 mM Mg^{2+}
Lane 8: λ *Pst*I

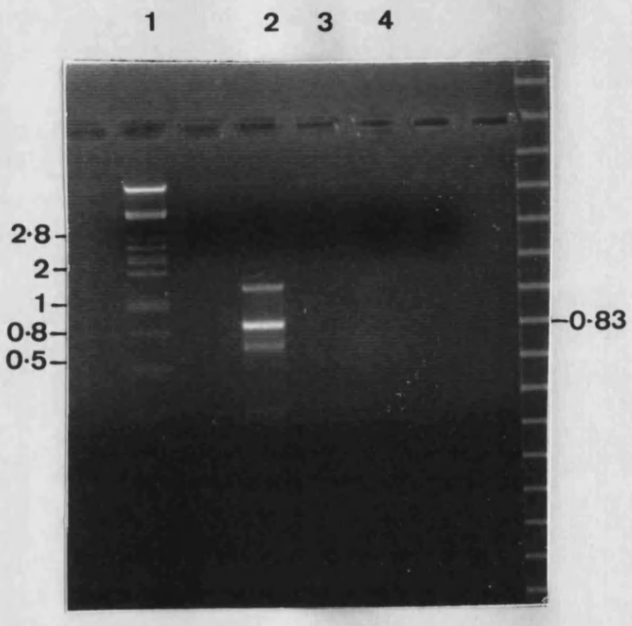

Figure 6.5 PCR amplification of genomic DNA using 29 base N-terminal and 26 base consesus primers, annealing at 55°C

Lane 1: λ *Pst*I

Lane 2: 1.5 mM Mg^{2+}

Lane 3: 1.5 mM Mg^{2+} 29 base N-terminal primer control

Lane 4: 1.5 mM Mg^{2+} 26 base consensus primer control

The control reactions contained only the primer indicated.

To confirm whether this product was correct, both ends were sequenced by the method described below.

A 10 µl aliquot of a 100 µl PCR reaction (1.5 mM Mg^{2+}) was run on a 1% agarose gel, the 830 bp DNA band excised, purified using GeneClean and dissolved in 50 µl sterile Milli Q water. This product was re-amplified at the same Mg^{2+} concentration, annealing at 60 or 65°C, using 0.5 µl as template (Figure 6.6).

Re-amplification at 60°C produced the highest yield of the single band. Asymmetric PCR was carried out at this annealing temperature to produce single-stranded DNA for sequencing.

6.2.3 Asymmetric PCR amplification of 830 bp fragment

Asymmetric PCR was carried out as described in Section 2.3.15.2, using 0.5 µl of the purified 830 bp band as template, at an annealing temperature of 60°C and at an Mg^{2+} concentration of 1.5 mM. Thirty cycles were performed, melting at 94°C (1 minute), annealing at 55°C (1 minute 30 seconds) and extending at 72°C (2 minutes). One reaction contained 50 pmoles of the N-terminal primer and 1 pmole of the consensus primer (to produce an excess of the sense strand), while a second reaction contained 50 pmoles of the consensus primer and 1 pmole of the N-terminal primer (to produce an excess of the anti-sense strand).

The single-stranded DNA produced by both of these reactions was sequenced.

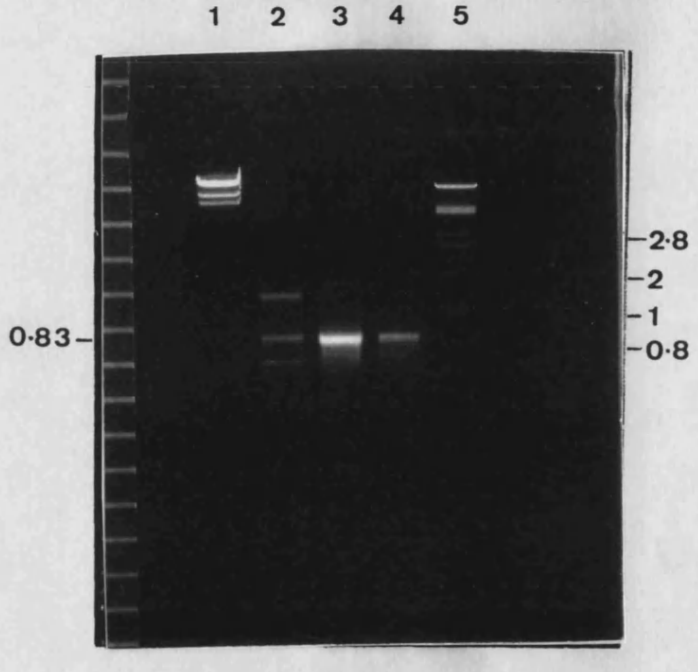

Figure 6.6 PCR amplification of 830 bp band using 29 base N-terminal and 26 base consensus primers

Lane 1: λ *Hin*dIII

Lane 2: Amplification from genomic DNA, 1.5 mM Mg^{2+}, annealing 55°C

Lane 3: Re-amplification of 830 bp band, 1.5 mM Mg^{2+}, annealing 60°C

Lane 4: Re-amplification of 830 bp band, 1.5 mM Mg^{2+}, annealing 65°C

Lane 5: λ *Pst*I

6.2.4 Sequencing of asymmetric PCR products

Sequencing was carried out by the dideoxy method described in Section 2.3.16, omitting the denaturation step. The sense strand was sequenced with 1 pmole of the consensus primer, and the anti-sense strand with 1 pmole of the N-terminal primer. The Sequenase label mix was diluted 10 fold and 2 minute extension times were used in order to sequence close to the primers.

The sequence obtained by extension from the N-terminal primer was translated in three reading frames, while the sequence obtained by extension from the consensus primer was reversed and complemented (to read the sense strand) and then translated in three reading frames (Figure 6.7). No translated sequence showed any homology with any previously determined citrate synthase sequence, confirming this product to be an artefact. Both sequences were then translated in all reading frames and used to search the GenBank sequence database via the FASTA program from the GCG program package. No significant identity with any sequence in the database was observed.

6.3 PCR amplification of a 960 bp sequence from genomic DNA using determined N-terminal and C-terminal amino acid sequences for primer design

Once the internal citrate synthase protein sequence described in Chapter Five became available, improved PCR reactions were undertaken, using these data.

Sequence of 830 bp PCR product obtained with N-terminal primer:

```
 S  T  L  *  E  V  S  R  I  S  R  P  R  F  R  R  A  P  S  G  E
  R  R  C  K  R  C  R  E  S  L  D  R  D  F  A  A  L  H  Q  A  N
   D  A  V  R  G  V  E  N  L  S  T  A  I  S  P  R  S  I  R  R  M
TCGACGCTGTAAGAGGTGTCGAGAATCTCTCGACCGCGATTTCGCCGCGCTCCATCAGGCGAA
        10         20         30         40         50         60
AGCTGCGACATTCTCCACAGCTCTTAGAGAGCTGGCGCTAAAGCGGCGCGAGGTAGTCCGCTT
 R  R  Q  L  L  H  R  S  D  R  S  R  S  K  A  A  S  W  *  A  F
  S  A  T  L  P  T  S  F  R  E  V  A  I  E  G  R  E  M  L  R  I
   V  S  Y  S  T  D  L  I  E  R  G  R  N  R  R  A  G  D  P  S  H
```

Sequence of 830 bp PCR product obtained with consensus primer:

```
 A  E  Q  P  R  I  P  A  R  T  A  *  R  Q  K  L  F  Y  P  C
  L  N  S  P  A  F  R  R  E  P  L  D  D  R  N  Y  F  T  R  A
   *  T  A  P  H  S  G  A  N  R  L  T  T  E  T  I  L  P  V  P
GCTGAACAGCCCCGCATTCCGGCGCGAACCGCTTGACGACAGAAACTATTTTACCCGTGC
        10         20         30         40         50         60
CGACTTGTCGGGGCGTAAGGCCGCGCTTGGCGAACTGCTGTCTTTGATAAAATGGGCACG
 S  F  L  G  A  N  R  R  S  G  S  S  S  L  F  *  K  V  R  A
  Q  V  A  G  C  E  P  A  F  R  K  V  V  S  V  I  K  G  T  G
   S  C  G  R  M  G  A  R  V  A  Q  R  C  F  S  N  *  G  H  R

 R  L  D  V  R  S  Y  Q  T  N  H  G  D  Y  R  H  L  S  D  E
  D  *  M  *  G  R  T  K  Q  I  M  E  I  T  D  I  S  A  T  K
   I  R  C  E  V  V  P  N  K  S  W  R  L  Q  T  S  Q  R  R  K
CGATTAGATGTGAGGTCGTACCAAACAAATCATGGAGATTACAGACATCTCAGCGACGAA
        70         80         90        100        110        120
GCTAATCTACACTCCAGCATGGTTTGTTTAGTACCTCTAATGTCTGTAGAGTCGCTGCTT
 S  *  I  H  P  R  V  L  C  I  M  S  I  V  S  M  E  A  V  F
  I  L  H  S  T  T  G  F  L  D  H  L  N  C  V  D  *  R  R  F
   N  S  T  L  D  Y  W  V  F  *  P  S  *  L  C  R  L  S  S  F

 N  Q  Q  R  V  V  G  R  V  H  R  V  P  A  R  H
  I  S  N  E  S  W  G  E  F  I  E  F  P  L  V  T
   S  A  T  S  R  G  A  S  S  S  S  S  R  S  S  P
AATCAGCAACGAGTCGTGGGGCGAGTTCATCGAGTTCCCGCTCGTCACCA
       130        140        150        160
TTAGTCGTTGCTCAGCACCCCGCTCAAGTAGCTCAAGGGCGAGCAGTGGT
 I  L  L  S  D  H  P  S  N  M  S  N  G  S  T  V
  D  A  V  L  R  P  A  L  E  D  L  E  R  E  D  G
   *  C  R  T  T  P  R  T  *  R  T  G  A  R  *  W
```

Figure 6.7 Sequences of each end of the 830 bp PCR fragment

Single-stranded DNA was produced by asymmetric PCR and sequenced with the same primers as were used for the amplification.

6.3.1 Primer design

The sense primer used was the non-degenerate 29 base oligonucleotide made to the citrate synthase N-terminal amino acid sequence (described in Figure 4.16).

The anti-sense primer was designed from the amino acid sequence of the cyanogen bromide cleavage fragment described in Section 5.5.2, with reference to the codon-usage data in Figure 4.1. Its sequence is shown in Figure 6.8.

The expected length of the PCR product was approximately 980 bp.

6.3.2 PCR amplification with N-terminal and C-terminal primers

PCR reactions were carried out using 80 ng genomic DNA as template, 0.4 µM primers and 50 µM dNTPs. Thirty cycles were performed, melting at 94°C Sequenceof 830 bp PCR product obtained with N-terminal primer:(1 minute), annealing at 55°C (1 minute 30 seconds) and extending at 72°C (2 minutes). The magnesium concentration in the reactions was varied from 0.5 to 5 mM. Figure 6.9a shows examples of the PCR products, separated by electrophoresis on a 1% agarose gel. Control reactions performed at the same time indicate that at 55°C (and 20 mM Mg^{2+}) the N-terminal primer binds at more than one position, causing products to be formed when it is the sole primer present. These products were larger than the desired fragments and while undesirable, did not interfere with their detection at this point (Figure 6.9b). At an annealing temperature of 55°C and a Mg^{2+} concentration of 5 mM, a product of 960 bp is formed. This product was of the expected size, although the amplification was not very specific.

```
          I   P   I   D   L   Y   T   P   I   F   A   V   S   R   A   G

        ATA CCA ATA GAC CTA TAC AAA CCA ATA TTC GCA GTA TCA CGA GCA GGA
          C   G   C   T   G   T   G   G   C   T   G   G   G   G   G   G
          T   C   T       C       C   C   T       C   C   C   C   C   C
              T           T       T   T           T   T   T   T   T   T
                      TTA                                 AGC
                        G                                   T

Sense     ATC CCG ATC GAC CTC TAC ACG CCC ATC TTC GC

Oligo     TAG GGC TAG CTG GAG ATG TGC GGG TAG AAG CG -5'
              G
```

Figure 6.8 Sequence of anti-sense PCR primer designed from amino acid sequence

The amino acid sequence and potential codons are shown. The primer sequence is in **bold** and includes the degeneracy shown.

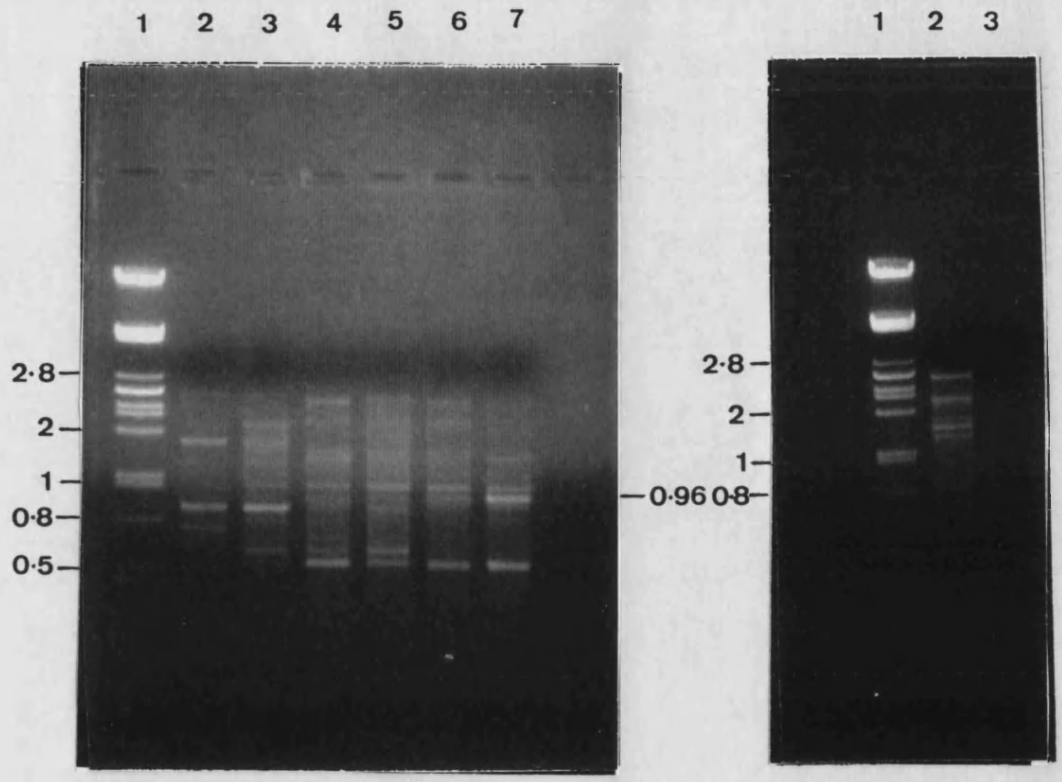

Figure 6.9a,b PCR amplification 960 bp band using N-terminal and C-terminal primers

a. Lane 1: λ *Pst*I
Lane 2: 0.5 mM Mg^{2+}
Lane 3: 1.0 mM Mg^{2+}
Lane 4: 1.5 mM Mg^{2+}
Lane 5: 2.0 mM Mg^{2+}
Lane 6: 3.0 mM Mg^{2+}
Lane 7: 5.0 mM Mg^{2+}

b. Lane 1: λ *Pst*I
Lane 2: 2 mM Mg^{2+}, N-terminal primer control
Lane 3: 2 mM Mg^{2+}, C-terminal primer control

As many PCR products were formed, the annealing temperature was raised to 60°C to reduce the level of non-specific primer binding. The result in Figure 6.10 demonstrates an improvement in the N-terminal primer control as products are no longer formed by this primer alone. The only product whose yield was improved by the higher annealing temperature is an 860 bp fragment evident in the 0.5-1.5 mM Mg^{2+} lanes. This is approximately 120 bp (corresponding to 40 residues) shorter than would be expected and unlikely to contain the desired sequence. This product was purified and re-amplified by asymmetric PCR as described in 6.2.3, in order to confirm this assertion; however, under these conditions the yield of single-stranded DNA was too low to permit sequencing.

The 960 bp product amplified at an annealing temperature of 55°C and 5 mM Mg^{2+} was investigated further. A 10 µl aliquot of a 100 µl PCR reaction (5 mM Mg^{2+}) was run on a 1% agarose gel, the 960 bp DNA band excised, purified using GeneClean and dissolved in 50 µl sterile Milli Q water. This product was re-amplified over a range of Mg^{2+} concentrations, annealing at 60°C, using 0.5 µl as template (Figure 6.11).

Asymmetric PCR was also carried out under the conditions described in 6.2.3 to confirm the presence of citrate synthase sequence. The yield of single-stranded DNA was too low to allow direct sequencing by this method.

6.4 Discussion

The current state of this project is that a 960 bp DNA fragment has been amplified from genomic DNA. Due to time constraints I was unable to confirm that this product contains a citrate synthase gene sequence. This is currently being undertaken by our research group by a route involving blunt-ended

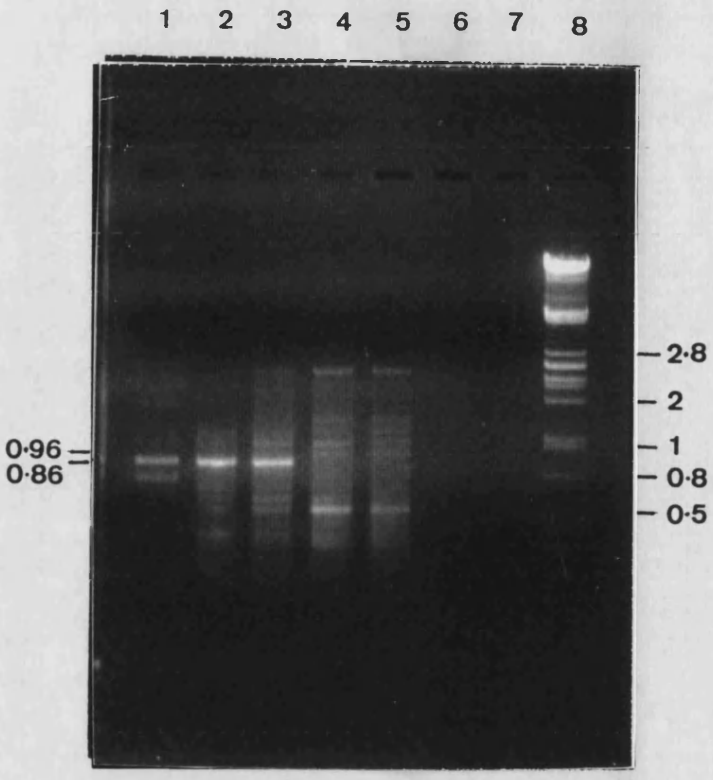

Figure 6.10 PCR amplification 960 bp band using N-terminal and C-terminal primers, annealing at 60°C

Lane 1: 1.0 mM Mg^{2+}
Lane 2: 1.5 mM Mg^{2+}
Lane 3: 2.0 mM Mg^{2+}
Lane 4: 3.0 mM Mg^{2+}
Lane 5: 5.0 mM Mg^{2+}
Lane 6: 2.0 mM Mg^{2+}, N-terminal primer control
Lane 7: 2.0 mM Mg^{2+}, C-terminal primer control
Lane 8: λ *Pst*I

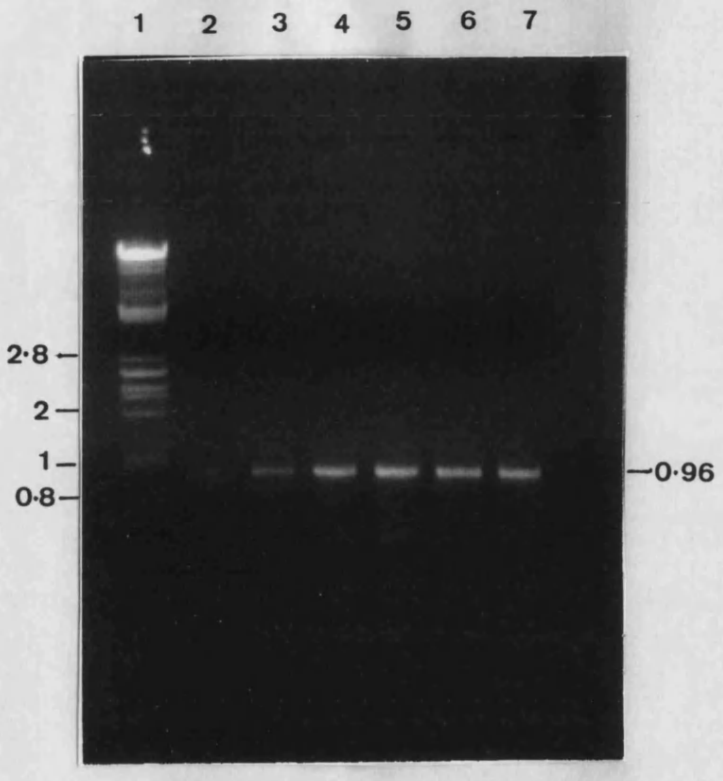

Figure 6.11 PCR amplification 960 bp band using N-terminal and C-terminal primers, annealing at 60°C

Lane 1: λ *Pst*I
Lane 2: 0.5 mM Mg^{2+}
Lane 3: 1.0 mM Mg^{2+}
Lane 4: 1.5 mM Mg^{2+}
Lane 5: 2.0 mM Mg^{2+}
Lane 6: 3.0 mM Mg^{2+}
Lane 7: 5.0 mM Mg^{2+}

cloning of the re-amplified product into a pUC vector. This will allow normal double-stranded DNA sequencing methods to be used. The 860 bp PCR product is shorter than one would expect for a section of a citrate synthase gene amplified between the N-terminal and C-terminal primers. The expected length of a product is some 120 bp greater, corresponding to about 40 amino acids in the protein. This difference is too small to dismiss this product as an artefact and should therefore be sequenced and retained or discarded on the basis of sequence data.

Once a gene sequence is obtained, perfectly matched oligonucleotide probes will be synthesised and used to probe restriction digested genomic DNA on Southern blots. By making probes close to both the N-terminal and C-terminal ends of the gene, a restriction fragment that has a high probability of containing the entire gene and flanking regions can be identified. Such a fragment would be at least 2 kb in length and hybridise to both probes.

Chapter 7

Analysis of 939 bp open reading frame within the 2.9 kb *Pst*I / *Eco*RI clone

7.1 Introduction

Chapter Four describes the screening of genomic DNA restriction fragments with a 44 base oligonucleotide probe designed from the N-terminal sequence of *Hf.volcanii* citrate synthase. A 2.9 kb *Pst*I / *Eco*RI fragment was isolated and cloned using this probe. Sequence analysis of this clone revealed that it did not contain the citrate synthase gene, although a FASTA search (GCG program package) of the Swissprot protein database with the translated DNA sequence revealed a putative open reading frame of an unrelated gene. This chapter describes the sequencing and analysis of the 939 bp open reading frame.

7.2 Sequence analysis

Sequencing was carried out as described in Chapter Four using the protocols given in Section 2.3.16. As the full characterisation and expression of this gene was outside the scope of this project, the sequence was determined in a single direction only. Figure 7.1 shows the position of the gene within the 2.9 kb insert and indicates the regions sequenced.

The gene sequence and corresponding amino acid sequence are given in Figure 7.2.

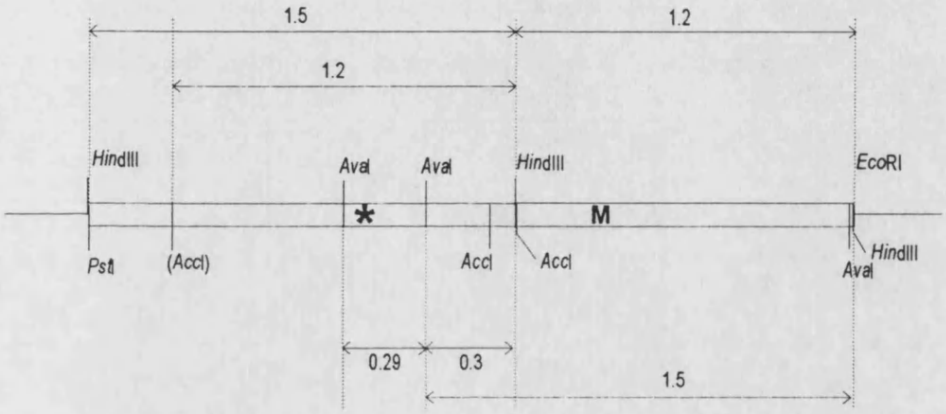

Figure 7.1 Position of the open reading frame within the 2.9 kb *Pst*I / *Eco*RI clone

The positions of the start codon (M) and stop codon (*) are indicated. The open reading frame is on the reverse strand.

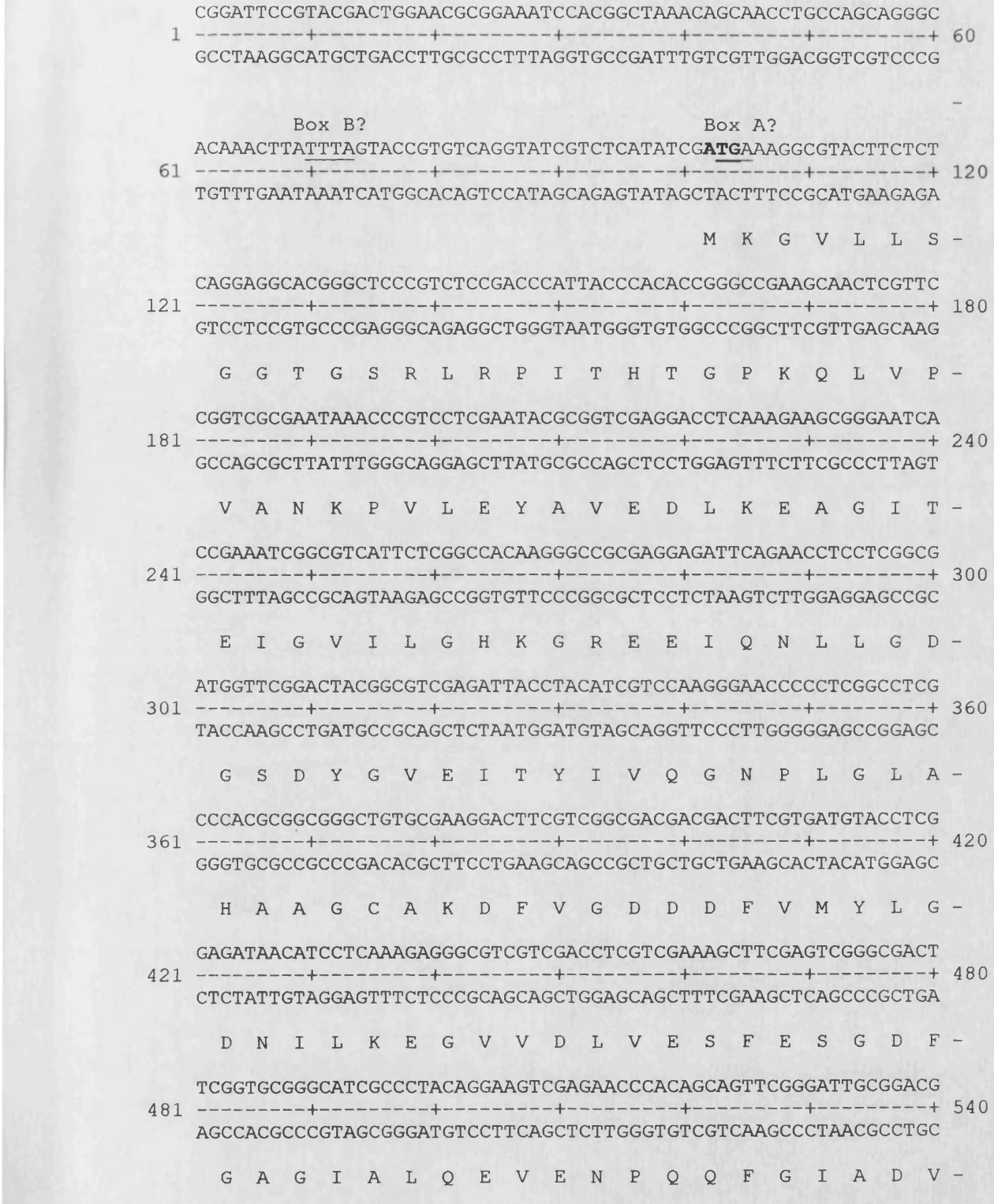

Figure 7.2 Sequence of open reading frame and corresponding amino acid sequence

The start codon is shown in bold. Putative promotor regions (Box A and B) are underlined.

```
    TCGACGACCAGGGTAACGTCACCCAACTCATCGAGAAGCCGGACGAGCCGCCGACGAATC
541 ---------+---------+---------+---------+---------+---------+ 600
    AGCTGCTGGTCCCATTGCAGTGGGTTGAGTAGCTCTTCGGCCTGCTCGGCGGCTGCTTAG

      D  D  Q  G  N  V  T  Q  L  I  E  K  P  D  E  P  P  T  N  L -

    TCGCACTCATCGGGATGTACGTCTTCTCGCCCGCGGTCTTCGACGCCATCGAGCAACTCG
601 ---------+---------+---------+---------+---------+---------+ 660
    AGCGTGAGTAGCCCTACATGCAGAAGAGCGGGCGCCAGAAGCTGCGGTAGCTCGTTGAGC

      A  L  I  G  M  Y  V  F  S  P  A  V  F  D  A  I  E  Q  L  E -

    AACCCTCGTGGCGCGGCGAACTCGAAATCACGGACGCGATTCAGTCACTCCTCGAAGACG
661 ---------+---------+---------+---------+---------+---------+ 720
    TTGGGAGCACCGCGCCGCTTGAGCTTTAGTGCCTGCGCTAAGTCAGTGAGGAGCTTCTGC

      P  S  W  R  G  E  L  E  I  T  D  A  I  Q  S  L  L  E  D  G -

    GCTACGCCATCGACTCGCACGTCGTCGAAGGCTGGTGGAAAGACACCGGCAAGCCCGAAG
721 ---------+---------+---------+---------+---------+---------+ 780
    CGATGCGGTAGCTGAGCGTGCAGCAGCTTCCGACCACCTTTCTGTGGCCGTTCGGGCTTC

      Y  A  I  D  S  H  V  V  E  G  W  W  K  D  T  G  K  P  E  D -

    ACATCCTCGAGGCGAACCAACTCGTCCTCGAAGACAAGTCGCTCAAAAAGCGGGGCACCG
781 ---------+---------+---------+---------+---------+---------+ 840
    TGTAGGAGCTCCGCTTGGTTGAGCAGGAGCTTCTGTTCAGCGAGTTTTTCGCCCCGTGGC

      I  L  E  A  N  Q  L  V  L  E  D  K  S  L  K  K  R  G  T  V -

    TCAGCGACGATGAGACCGTCGATGGACGTATCGAACTCGCGGAGTCAGCGACCATCGAAG
841 ---------+---------+---------+---------+---------+---------+ 900
    AGTCGCTGCTACTCTGGCAGCTACCTGCATAGCTTGAGCGCCTCAGTCGCTGGTAGCTTC

      S  D  D  E  T  V  D  G  R  I  E  L  A  E  S  A  T  I  E  D -

    ACGGCGCAGTCGTCCGTGGCCCGGTCTCTATCGCGGATGGCGCGGTCATCAAATCGGGCA
901 ---------+---------+---------+---------+---------+---------+ 960
    TGCCGCGTCAGCAGGCACCGGGCCAGAGATAGCGCCTACCGCGCCAGTAGTTTAGCCCGT

      G  A  V  V  R  G  P  V  S  I  A  D  G  A  V  I  K  S  G  T -

    CCTACGTCGGCCCGTACACGTTTCGGTCCGAACTCGACACTCGAAGGCGTCCACATCGAG
961 ---------+---------+---------+---------+---------+---------+ 1020
     GGATGCAGCCGGGCATGTGCAAAGCCAGGCTTGAGCTGTGAGCTTCCGCAGGTGTAGCTC

      Y  V  G  P  Y  T  F  R  S  E  L  D  T  R  R  R  P  H  R  E -

    AACTCGGTCGTCATCGGTGAGTCGAGCATCAACACCTCCGGCCGCATCGTTGACAGCCTC
1021 ---------+---------+---------+---------+---------+---------+ 1080
    TTGAGCCAGCAGTAGCCACTCAGCTCGTAGTTGTGGAGGCCGGCGTAGCAACTGTCGGAG

      L  G  R  H  R  *                                           -
```

Figure 7.2 Continued

```
     CTCGGGAAGGGTCGCGAACATCGGCAGTGCGGACGACTTCCTCCCGGAGGGTCGTCGCCT
1081 ---------+---------+---------+---------+---------+---------+ 1140
     GAGCCCTTCCCAGCGCTTGTAGCCGTCACGCCTGCTGAAGGAGGGCCTCCCAGCAGCGGA

                                                                  -

     CGTCGTTGGCGAGGGCTCCCAACTGAAACTCTAACCATGCACATCATCACCCACGCCTGT
1141 ---------+---------+---------+---------+---------+---------+ 1200
     GCAGCAACCGCTCCCGAGGGTTGACTTTGAGATTGGTACGTGTAGTAGTGGGTGCGGACA

                                                                  -

     ACGCAGTGCGGAACCGTCGTCTCCGCGAACGAACTCGAATCGAACCGCGTGATGAAGTGC
1201 ---------+---------+---------+---------+---------+---------+ 1260
     TGCGTCACGCCTTGGCAGCAGAGGCGCTTGCTTGAGCTTAGCTTGGCGCACTACTTCACG

                                                                  -

     CCCGGA
1261 ------ 1266
     GGGCCT

            -
```

Figure 7.2 Continued

7.3 Database search results

The Genbank and EMBL databases were searched with the complete translated sequence of the open reading frame. The results of the search are given in Table 7.1 below.

Sequence	Organism	% identity with ORF	GenBank Accession No.	Reference
strD (putative glucose-1-phosphate thymidylyltransferase)	*Streptomyces griseus*	47.1% in 293 aa overlap	Y00459	Distler *et al.* (1987)
Capsule gene complex UDP-glucose-4-epimerase	*Neisseria meningitidis*	36.9% in 244 aa overlap	L09188	Hammerschmidt *et al.* (1992)
Genomic region 84.5 to 86.5	*Escherichia coli*	36.3% identity in 237 aa overlap	M87049	Daniels *et al.* (1992)
Genomic region 325 to 333	*Bacillus subtilis*	40.6% in 244 aa overlap	X73124	Glaser *et al.* (1993)
rfbG (putative glucose-1-phosphate thymidylyltransferase)	*Yersinia enterocolitica*	33.3% in 273 aa overlap	Z18920	Zhang *et al.* (1993)
rfbB (dTDP-glucose 4,6-dehydratase)	*Shigella flexneri*	36.9% in 236 aa overlap	L14842	Rajakumar *et al.* (1994)
rfbC or rfbD (dTDP-6-deoxy-D-glucose-3,5 epimerase OR dTDP-6-deoxy-L-mannose dehydrogenase)	*Salmonella typhimurium*	36.9% in 236 aa overlap	X56793	Jiang *et al.* (1991)
rfbA (glucose-1-phosphate thymidylyltransferase)	*Salmonella enterica*	36.9% in 236 aa overlap	X60665	Wang *et al.* (1992)
rfbA (glucose-1-phosphate thymidylyltransferase)	*Azorhizobium caulinodans*	40.4% in 208 aa overlap	Z22611	Goethals *et al.* (1993)
rfbA (putative glucose-1-phosphate thymidylyltransferase)	*Xanthomonas campestris*	38.6% in 236 aa overlap	L23941	Koeplin *et al.* (1993)

```
                1                                                    50
Hf.volcanii ORF ..... ..... ..... ..... mKgvlLsGGt
      S.griseus ..... ..... ..... ..... mKalvLaGGt
         E.coli ..... ..... ..... ..... mKgimLaGGs
 N.meningitidis vakcierqls fgtfrywkiv frrhpdamps enpsqrkkek mKgiiLaGGs
     B.subtilis ..... ..... ..... ..... vKgviLaGGn
      Consensus ---------- ---------- ---------- ---------- -K---L-GG-

               51                                                   100
Hf.volcanii ORF GsRLrPiTht gpKqLvPVan kPvleYaved lkeAGItEIg vIlghkgree
      S.griseus GtRLrPiTht saKqLvPVan kPvlfYglea iraAGIiDVg iVvgdta.de
         E.coli GtRLhPiTrg vsKqLlPIyd kPmiyYplsv lmlAGIrEIl iIttpedkgy
 N.meningitidis GtRLyPiTrg vsKqLlPVyd kPmiyYplsv lmlAGIrDIl vItapednas
     B.subtilis GsRLmPlTka vnKhLlPVgp yPmiyWsimk lqeAGIkDIl lIsqkehmpq
      Consensus G-RL-P-T-- --K-L-PV-- -P---Y---- ---AGI-DI- -I--------

               101                                                  150
Hf.volcanii ORF iqnllGdgsd YGveitYivQ gnplGlahaa gcakdFvgdD dfvMyLGDNi
      S.griseus ivaavGdgsr FGlkvsYipQ skplGlahcv lisrdFlgeD dfiMyLGDNf
         E.coli fqrllGdvge FGiqleYaeQ pspdGlaqaf iigetFlngE pscLvLGDNi
 N.meningitidis fkrllGdgsd LGisisYavQ pspdGlaqaf iigeeFignD nvcLvLGDNi
     B.subtilis fykllGngee LGvtitYqvQ paasGisdgl syakrFtkkE sfiLlLGDNi
      Consensus -----G---- -G----Y--Q ----G----- -----F---D ---L-LGDN-

               151                                                  200
Hf.volcanii ORF lke.gv.vlv esfesgdfga gialqeVenP qqFGiaDvdD qg.nvtqliE
      S.griseus vvg.vvedsv refraarpda hlmltrVpeP rsFGvaElsD sg.qvlgleE
         E.coli ffgqgfspkl rhvaartegp .vfgyqVmdP erFGxvEfdD .nfraisleE
 N.meningitidis fygqsftqtl kqaarqthga tvfayqVknP erFGvvEfnE .nfravsieE
     B.subtilis f.edslkpyt erfeqqgkga kvllkeVddP erFGiaEidE knkrirsiiE
      Consensus ---------- ---------- ------V--P --FG--E--D ---------E

               201                                                  250
Hf.volcanii ORF KPdePptnla liGmYvFspa vfdaieqleP SwRGELEITd aiqsllEdgy
      S.griseus KPahPksdla lvGvYlFspa iheavaaitP SwRGELEITd avqwliDagr
         E.coli KPkqPksnwa vtGlYfYdsk vveyakqvkP SeRGELEITs inqmylEagn
 N.meningitidis KPqrPksdwa vtGlYfYdnr avefakqikP SaRGELEITd lnrmylEdgs
     B.subtilis KPeqPptnlc vtGiYmYdae vfsyieqisP SkRGELEITd vnnlyiEnsq
      Consensus KP--P----- --G-Y-Y--- ---------P S-RGELEIT- ------E---

               251                                                  300
Hf.volcanii ORF aidshvveGW .WkDtGkped ileanqLVle dkslkkrgtv sddetvdgri
      S.griseus dvrstvisGY .WkDtGnvtd mlevnrLVle tteprcdglv dersdligrv
         E.coli ltvellgrGF aWlDtGthds lieastFV.q tvekrqgfki acleeiawrn
 N.meningitidis lsvqilgrGF aWlDtGthes lheaasFV.q tvqniqdlqi acleeiawrn
     B.subtilis ltydvls.GW .WvDaGthes lylasqLVhq alrkgqdek* ..........
      Consensus --------G- -W-D-G---- ------LV-- ---------- ----------
```

Figure 7.3 Alignment of the *Hf.volcanii* open reading frame (ORF) with the four sequences of highest identity from the database search

The uppercase letters denote conserved or conservatively replaced residues. The alignment was produced using PILEUP. Refer to Table 7.1 for the functions of these genes.

```
                301                                                  350
Hf.volcanii ORF elaesatied gavvrgpvsi adgaviksgt yvgpytfrse ..........
     S.griseus  lveegaevrn srvmgptvig agtrvtns.. yvgpftslae dcvvedseve
       E.coli   gwlddegvkr aasslaktgy gqyllellra rprqy*.... ..........
 N.meningitidis gwlekkdvet rakplektay gqyllhligk *......... ..........
    B.subtilis  .......... .......... .......... .......... ..........
      Consensus ---------- ---------- ---------- ---------- ----------

                351                                                  400
Hf.volcanii ORF .......... ldtrrrphre lgrhr*.... .......... ..........
     S.griseus  fsivlrgasi sgvrrieasl igrhvqvtsa pevphanrlv lgdhsraqis
       E.coli   .......... .......... .......... .......... ..........
 N.meningitidis .......... .......... .......... .......... ..........
    B.subtilis  .......... .......... .......... .......... ..........
      Consensus ---------- ---------- ---------- ---------- ----------

                401
Hf.volcanii ORF ..
     S.griseus  s*
       E.coli   ..
 N.meningitidis ..
    B.subtilis  ..
      Consensus --
```

Figure 7.3 Continued

13%. This is consistent with the theory that halophilic proteins contain acidic residues that stabilise them under high salt conditions, as described in Section 1.6.

Of the sequences given in Table 7.1, two are of unknown function (*B.subtilis* genomic region 325-333 and *E.coli* genomic 84.5-86.5), one is a UDP-glucose-4-epimerase, one a dTDP-glucose 4,6-dehydratase and one either a dTDP-6-deoxy-D-glucose-3,5 epimerase or a dTDP-6-deoxy-L-mannose dehydrogenase. The remainder are putative (*S.griseus*, *Y.enetrocolitica* and *X.campestris*) or confirmed (*S.enterica*) glucose-1-phosphate thymidylyltransferases.

The *S.griseus* glucose-1-phosphate thymidylyltransferase is proposed to be involved early in the biosynthetic pathway of streptomycin (Distler *et al*, 1987, Distler *et al.*, 1992). With the exception of the *S.flexneri* dTDP-glucose 4,6-dehydratase, roles have been proposed for all of the other activities in the biosynthesis of lipopolysaccharides. In *Y.enterocolitica* and the *Salmonella* species the lipopolysaccharide in question is a surface antigen, while the *A.caulinodans* gene is part of a lipopolysaccharide operon involved in nodule formation.

The *Hf.volcanii* open reading frame may code for an enzyme that has glucose-1-phosphate thymidylyl or urdiylyl transferase activity, activating a carbohydrate moiety by nucleotidylation prior to its polymerisation into a polysaccharide. Alternatively the gene product may accept a nucleotidylated sugar as a substrate and act as an epimerase or dehydratase.

If the *Hf.volcanii* open reading frame codes for a sugar-activating enzyme, it is likely to have a fundamental role in polysaccharide biosynthesis. It has been proposed that nucleotide-activated sugars (UDP-glucose, UDP-mannose and UDP-galactose) are added to an unidentified glycolipid precursor during the biosynthesis of glycolipids in *Hf.volcanii*

(Kates, 1994). The biosynthesis of the *Hf.volcanii* S-layer glycoprotein also utilises nucleotide-activated sugars (Kandler & Konig, 1994).

Further study of this open reading frame requires confirmation of its sequence by sequencing both strands, followed by characterisation of the activity of the protein which it encodes. This could be approached by PCR amplification of the gene using primers containing restriction sites which would allow it to be cloned into an *E.coli* expression vector. The gene could then be expressed in *E.coli* under the control of a powerful, inducible promotor (such as *tac*) and the enzyme activated by addition of salt. If the expressed protein is found to be insoluble, solubility may be improved by growing the transformed *E.coli* at 30-32°C. Activation of the enzyme may not be achieved by the simple addition of KCl to a cell extract. It may be necessary to partially unfold the expressed enzyme and allow it to refold into an active conformation by addition of guanidine hydrochloride followed by dialysis against a high KCl buffer, or by heating the inactive enzyme in a high KCl buffer and allowing it to cool.

The development of an assay to determine the activity of the expressed enzyme should be undertaken with care as the enzyme may be inactive due to incorrect folding. Adding 2 M KCl to a cell extract (produced by sonication) acts as an initial purification as many *E.coli* proteins are precipitated under these conditions. A subsequent dialysis step would then remove nucleotide triphosphates and nucleotidylated sugars which could interfere with assays.

Simple spectrophotometric assays for glucose-1-phopshate thymidylyltransferase rely on addition of pyrophosphate and dTDP-glucose as substrates, catalysing the formation of glucose-1-phosphate, which is in turn converted to glucose-6-phosphate by phosphoglucomutase (in the presence of its activator, glucose-1,6-bisphosphate) which is coupled to glucose-6-phosphate dehydrogenase and the reduction of NAD is followed

spectrophotometrically (Kornfeld & Glaser, 1961). This type of coupled assay is unsuitable for use with a halophilic enzyme. An alternative system has been developed using HPLC analysis of aliquots taken from a reaction between glucose-1-phosphate and dTTP, forming dTDP-glucose and pyrophosphate (Lindquist *et al.*, 1993). This method is also unsuitable as it relies on a non-halophilic pyrophosphatase to affect the equilibrium of the reaction.

A possible method for a halophilic enzyme would be to incubate the cell extract (prepared as described above) with a radiolabelled sugar-phosphate and dTDP or UDP, stop the reaction and purify the nucleotide diphosphate sugar by adsorption to charcoal, eluting with a solution of ammonia in ethanol/water (Okazaki *et al.*, 1961). The incorporation of radiolabel could be measured for a variety of different combinations of sugar-phosphates and dTDP or UDP, to determine the preferred substrate.

An epimerase activity may be detected by incubation of the enzyme with a potential substrate, followed by purificaction of the substrate-product mixture by adsorption as above. Analysis by HPLC of aliquots taken at various time points during the incubation would reveal any reaction that had taken place.

Chapter 8

Conclusions and suggested further work

The strategy of purifying a protein, determining its N-terminal amino acid sequence and using a corresponding synthetic oligonucleotide probe to screen for its gene is well established. In our laboratory this route has been successfully taken in cloning the citrate synthase gene (Sutherland *et al.*, 1990) and the glucose dehydrogenase gene (Bright *et al.*, 1991) from the Archaeon *Thermoplasma acidophilum.* In each case both long, unique probes and short, degenerate probes were used, the citrate synthase gene being identified using the former type of probe and the glucose dehydrogenase with the latter. Both of these approaches to the design of probes were therefore considered appropriate for this project.

Short, degenerate probes designed to *Hf.volcanii* citrate synthase amino acid sequence were found to lack the required specificity when hybridising to genomic DNA. Longer, unique probes appear to offer a greater chance of success in following up this work. The marked preference for certain codons in the *Hf.volcanii* genome makes their deign easier.

Trieselmann and Charlebois (1992) have shown by cDNA hybridisation to genomic DNA that regions of the relatively A-T rich FII DNA of the *Hf.volcanii* genome are transcriptionally active. There may be protein coding genes within these areas whose codon preference differs from that described in Figure 4.1. Probes designed using codon preference data from G-C rich regions would be less likely to hybridise to such genes. This may be relevant to our search for the citrate

synthase gene. A comprehensive survey of codon preference in halophilic Archaea has recently been carried out (Soppa, 1993). The majority of the genes surveyed possessed the same G-C biased codon preference, including those used to compile Figure 4.1. To date, only the gas vacuole genes of the *Halobacteria* are known to have a significantly different (relatively A-T) rich codon preference. On balance it still appears valid to have designed the probes to the G-C rich codon preference.

The next step to be taken with this project should be the cloning and sequencing of the 960 bp PCR product described in Section 6.3.2. A pair of 20-30 base oligonucleotides designed to match the ends of a correctly amplified product could be used to probe restriction-digested genomic DNA. A fragment of 2 to 3 kb to which both probes bound would be highly likely to contain the entire gene. Should both of the PCR products described in Section 6.3.2 prove to be artefacts, the 29 base PCR primer designed from the amino acid sequence at the C-terminal end of the protein has the potential to be used as a probe.

At the outset of this project it was our intention eventually to express the *Hf.volcanii* citrate synthase in *E.coli,* prepare crude cell extracts by sonication and use the addition of salt to both activate the *Hf.volcanii* enzyme and concommitantly to precipitate many of the contaminating *E.coli* proteins. Cendrin *et al.* (1993) have since demonstrated that it is possible to achieve high level expression of a halophilic malate dehydrogenase in a soluble form in *E.coli.* The enzyme was fully activated by the addition of NaCl to a concentration of 3 M. The *Hf. volcanii* dihydrofolate reductase has also been expressed in *E.coli*, but in an insoluble form that was activated by denaturing in guanidine hydrochloride, followed by dilution into high salt solutions.

If the *Hf.volcanii* citrate synthase is found to express as an incorrectly folded polypeptide that resists folding upon the addition of either high salt or guanidine, the option of expression in *Hf.volcanii* is still open. *Hf.volcanii* is a fast growing organism that reaches a high density in culture and an effective means of enzyme purification is available in the Red A affinity column. Expression in *Hf.volcanii* has the advantage that the polypeptide is likely to fold correctly, but has the drawback that expression would be from its own promotor rather than the highly efficient, inducible promotors (such as *tac*) available in *E.coli* expression systems.

In order to study the basis of halophilicity by site-directed mutagenesis one needs detailed knowledge of the structure of the enzyme under investigation. A crystal structure has yet to be determined for a halophilic enzyme, so until such time mutagenesis studies must rely upon predicted structural models, such as that for the *Hf.volcanii* dihydrofolate reductase (Bohm & Jaenicke, 1994). The crystal structure of the *Tp.acidophilum* citrate synthase has now been determined (Russell *et al.*, 1993) and may be a useful point from which to start homology modelling the *Hf.volcanii* enzyme once its sequence is known, while crystallisation trials are started. As citrate synthases are composed almost entirely of α-helices it will be interesting to see if any in the *Hf.volcanii* enzyme have the "halophilic helix motif" (negative charge clusters at the N-terminal end of a helix) observed in the dihydrofolate reductase model.

The sequencing of the *Hf.volcanii* citrate synthase will provide a representative of the halophiles for phylogenetic analysis, a group so far absent from the phylogenetic tree of citrate synthases. The N-terminal amino acid sequence of the *Hf.volcanii* citrate synthase indicates that the enzyme is, like that of *Tp.acidophilum* and

B.coagulans, approximately 40 residues shorter than other citrate synthases. Two other Bacillus sequences have since been determined (A.L. Sonenshein, Tufts Univ. School of Medicine, Boston, USA, Unpublished), both of which are of the shorter type. This points to a relationship between the citrate synthases of the Archaea and the Gram positive Bacteria. The Gram positive *Mycobacterium smegmatis* also groups with the Archaea and *Bacillus* species on phylogenetic analysis (Taylor, G., Hough, D.W. and Danson, M.J., Unpublished data), although it does not lack the 40 or so N-terminal residues.

It is important in defining phylogenetic relationships between organisms to take a consensus from the phylogenetic relationships between as many of their genes as possible. Not all gene sequences will represent their parent organism in the same fashion as all have developed under slightly different conditions. It may be argued that all genes are valuable phylogenetically; however, our present techniques of analysis lack the sophistication to unlock the information within many sequences. Citrate synthase is a widespread central metabolic enzyme, functionally conserved in all organsims that possess it. It is a valuable addition to the growing range of enzymes whose gene sequences are a source of accessible phylogenetic information.

During the course of this project an open reading frame was idenfied within a cloned fragment of *Hf.volcanii* genomic DNA. It has been shown in Chapter Seven that this sequence bears significant homology with several enzymes from bacterial sources involved in the metabolism of nucleotide-activated hexose sugars. Such enzymes have been implicated in the biosynthesis of lipopolysaccharides and S-layer glycoproteins in halophilic Archaea. This is a subject area that has not yet been intensively studied at the genetic level and no gene coding for any of these activities has yet been cloned. Chapter Seven described

possible approaches to characterise the function of the *Hf.volcanii* open reading frame.

If this sequence is a gene involved in polysaccharide metabolism it may provide a basis for further investigation into these pathways in the halobacteria, particularly if other genes of related function in the pathway are clustered with this one, as is the case in the *Salmonella* and *Yersinia* species studied to date. The sequencing of such a gene cluster, while a major project, would provide data from which a putative metabolic pathway could be deduced.

Appendix

Purification of citrate synthase

Citrate synthase from *Hf.volcanii* was purified in our laboratory by Dr. M.J. Bonete, University of Alicante, Spain using a scheme based on that used to purify the *H.halobium* citrate synthase (James, 1990).

The purification is summarised in the table below.

Step	Protein mg	CS Activity U	Sp. Activity U mg^{-1}	Purification	Yield %
Cell-extract	2105	78.3	0.037	1	100
DEAE-Sepharose	350	64.7	0.185	5	83
Hydroxylapatite	53.5	58.5	1.096	30	75
Mono-Q	7.8	37.3	4.78	129	48
Superdex-200	0.53	10.0	18.88	510	13
Superdex-200	0.13	3.4	25.8	697	4

SDS-polyacrylamide gel electrophoresis and Western blotting

The dried citrate synthase was dissolved in 40 µl Milli Q water. SDS-polyacrylamide gel electrophoresis was carried out on 2 µl aliquots using the Bio-Rad Protean II system. An initial gel was Coomassie stained to locate the protein bands, while subsequent ones were left unstained in order to carry out Western blots. The latter were used to optimise blotting

conditions. Figure 9.1 below shows a 0.5 µl aliquot of the purified citrate synthase run on a small (Mini-Atto system) 8% SDS polyacrylamide gel. Two major protein bands are visible; one of 45 kD and another of approximately 55 kD.

The remaining 30 µl were electrophoresed as above and blotted onto Immobilon P membrane under the optimised conditions. The membranes were stained and both major protein bands (45 kD and 55 kD) were cut out. The bands were sequenced directly from the membrane.

N-terminal sequencing

Both protein bands were sequenced using an Applied Biosystems 470A gas-phase Sequencer coupled to an Applied Biosystems 120 Phenylthiohydantoin analyser.

Sequences were obtained for both 45 kD and 55 kD bands. These are shown below, using single letter amino acid codes where X indicates an unknown residue and residues in parentheses are tentative assignments.

55 kD XXXXVIAAAYRTPFGKAA

45 kD (S)GELKRGLEGVLVAESKLSFIDGDAGQLVYXGYDIEDL

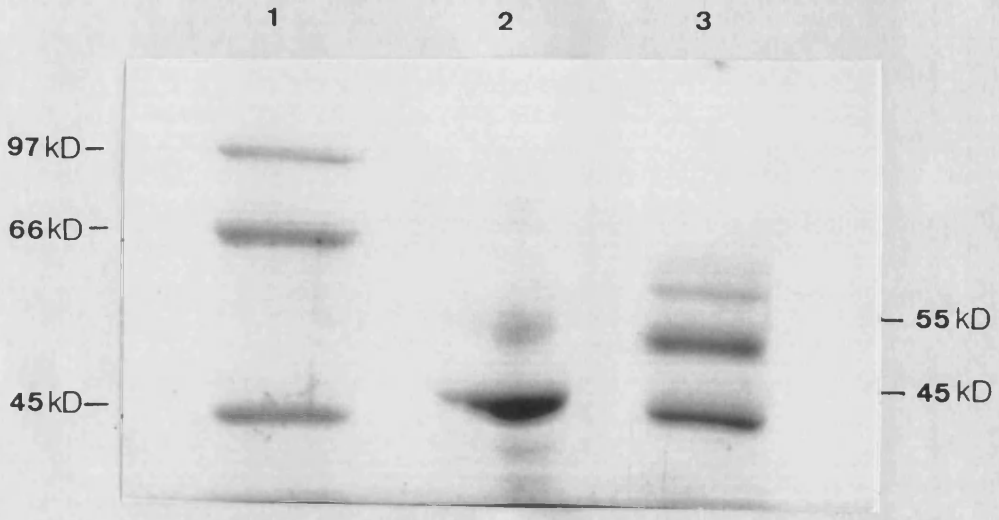

Figure 9.1 SDS-PAGE of purified *Hf.volcanii* citrate synthase

Lane 1: Molecular weight markers

Lane 2: Pig heart citrate synthase (4 µg)

Lane 3: *Hf.volcanii* citrate synthase (0.5 µl of 40 µl)

Halophilic proteins are often highly acidic and thus bind less of a cationic detergent, such as SDS, than a non-halophilic protein. This results in reduced mobility on SDS gels and consequent over-estimates of protein molecular weight. It was therefore not possible to distinguish which of the two major bands was citrate synthase by migration distance alone. Based on the subunit sizes observed for citrate synthases from other sources, the 45 kD band is closest to to the expected size of a citrate synthase monomer; however, the 55 kD band could be citrate synthase migrating slowly for the reasons described above. In the latter case the 45 kD band could either be a proteolytic degradation product of the larger band or a co-purified contaminant. It was therefore necessary to obtain N-terminal sequence for both of the bands.

The sequence of the 45 kD band was initially aligned by hand with a multiple sequence alignment of eight other citrate synthases. The sequences from both bands were used to search the Swissprot (1991) protein database, using the FASTA search program from the GCG program package. Only the 45 kD band showed significant homology with any of the sequences in the database. All of the best scores, with the exception of one, were obtained for matches to the N-termini of citrate synthases. The eight highest identities are shown in Table 9.1.

Sequence	Organism	Swissprot Accession No.	Reference
Citrate synthase	*Coxiella burnetii*	P18789	Heinzen *et al.* (1991)
Citrate synthase	*Pseudomonas aeruginosa*	P14165	Donald & Duckworth (1989)
Citrate synthase	*Acinetobacter anitratum*	P20902	Donald & Duckworth (1987)
Citrate synthase	*Rickettsia prowazekii*	P09948	Wood *et al.* (1987)
Citrate synthase	*Acetobacter aceti*	P20901	Fukaya *et al.* (1990)
DNA dependent RNA polymerase	*Sulfolobus acidocaldarius*	P11513	Puehler *et al.* (1989)
Citrate synthase	*Thermoplasma acidophilum*	P21553	Sutherland *et al.* (1990)
Citrate synthase	*Escherichia coli*	P00891	Bhayana & Duckworth (1984)

Table 9.1 Results of Swissprot database search

Alignments of available citrate synthase sequences show that the enzymes can be divided into two categories based on the presence or absence of a stretch of approximately 40 amino acids at the N-terminus (Muir *et al.*, 1994a). The gene sequences of the *Tp.acidophilum* (Sutherland *et al.*, 1991) and *Bacillus* sp. (Schendel *et al.*, 1992) citrate synthases, which are both of the shorter type, show that the absence of

these residues is not due to proteolytic cleavage, but is coded for at the DNA level.

Figure 9.2 shows the relevant part of a citrate synthase multiple sequence alignment created using the PILEUP program from the GCG package. The alignment was created using the full sequences and includes the *Hf.volcanii* 45 kD band sequence and the recently determined N-terminal sequences from *S.solfataricus* (Lill *et al.*, 1992) and *P.furiosus* (Muir *et al.*, 1994b) citrate synthases. It is evident that the *Hf.volcanii* N-terminal sequence aligns with the shorter N-termini and hence the enzyme belongs to the short category. It is unlikely that this observation is due to proteolytic cleavage of the *Hf.volcanii* N-terminus as the lack of these residues appears to be a feature of archaeal citrate synthases.

```
                    1                                                  50
      Chicken heart ...asstnlk dvlaalipke qariktfrqq hggtalgqit v.dms.....ygGmrgmkgl vyetsvldpd egi.rfrGfs ipecq
          Pig heart ...asstnlk diladlipke qariktfrqq hgntvvgqit v.dmm.....ygGmrgmkgl vyetsvldpd egi.rfrGys ipecq
Yeast (cytoplasm)   ...sqektlk erfseiypih aqdvrqfvke hgktkisdvl l.eqv.....ygGmrgipgs vwegsvldpe dgi.rfrGrt iadiq
   Yeast (mitoch)   ...aseqtlk erfaeiipak aqeikkfkke hgktvigevl leeqa.....ygGmrgikgl vwegsvldpe egi.rfrGrt ipeiq
      Tetrahymena   ....sqtnlk kviaeiipqk qaelkevkek ygdkvvgqyt v.kqv.....igGmrgmkgl msdlsrcdpy qgi.ifrGyt ipqlk
     Mycobacterium  .......apq qalrtaalrr lqpvrlrhlq lrraggesmt tateseaprihkGlagvvvd ttaiskvvpe tnsltyrGyp vqd.l
         Bacillus   .......... .......... .......... .......... ....vntnqfipGlegvias etkisfldtv nseivikGyd lla.l
     Acinetobacter  ....seatgk kavlhldgke .ielpiysgt lgpdvidvkd vlas.ghftfdpGfmatasc eskitfidgd kgillhrGyp idq.l
      Pseudomonas   .......adk kaqliiegsa pvelpvlsgt mgpdvvdvrg ltat.ghftfdpGfmstasc eskityidgd kgvllhrGyp ieq.l
      Escherichia   .......adt kakltlngdt aveldvlkgt lgqdvidirt lgsk.gvftfdpGftstasc eskitfidgd egillhrGfp idq.l
      Acetobacter   sasqkegkls tatisvdgks a.empvlsgt lgpdvidirk lpaqlgvftfdpGygetaac nskitfidgd kgvllhrGyp iaq.l
         Coxiella   .......snr kaklsfenqs .vefpiyspt lgkdvidvkt l.gnhgayaldvGfystaac eskitfidge kgillyrGyp idq.l
       Rickettsia   .tngnnnnle faelkirgk. lfklpilkas igkdvidisr vsaeadyftydpGfmstasc qstityidgd kgilwyrGyd ikd.l
     Thermoplasma   .......... .......... .......... .......... ....peteeiskGledvnik wtrlttidgn kgilrygGys vedii
         Haloferax                                                 SGELKRGLEGVLVA ESKLSFIDGD AGQLVYAGYD IED L
         Consensus  ---------- ---------- ---------- ---------- ------------G------- ---------- -------G-- -----
```

Figure 9.2 Multiple sequence alignment of citrate synthases showing the *Hf.volcanii* N-terminal sequence

The alignment was created using the PILEUP from the GCG package. The entire citrate synthase sequences were used to create the alignment and the N-terminal sequence was then added by hand.

References

Aitken, D.M. & Brown, A.D. (1969) *Biochim. Biophys. Acta* 177, 351-354

Alter, G.M., Casazza, J.P., Wang, Z., Nemeth, P., Srere, P. & Evans, C.T. (1990) *Biochemistry* 29, 7557-7563

Baldacci, G., Guinet, F., Tillit, J., Zaccai, G. & de Recondo, A. (1990) *Nucleic Acid Res.* 18 No.3, 507-511

Beanland, T.J. & Howe, C.J. (1992) *Comp. Biochem. Physiol.* 102B No.4, 643-659

Benachenhou-Lahfa, N., Forterre, P. & Labedan, B. (1993) *J. Mol. Evol.* 36, 335-346

Bhayana, V. & Duckworth H.W. (1984) Biochemistry 23, 2900-2905

Bonnete, F., Ebel, C., Zaccai, G. & Eisenberg H. (1993) *J. Chem. Soc. Faraday Trans.* 89 No.15, 2659-2666

Bohm, G. & Jaenicke, R. (1994) *Protein Eng.* 7, 213-220

Bradford, M.M. (1976) *Anal. Biochem.* 72, 248-254

Cendrin, F., Chroboczek, J., Zaccai, G., Eisenberg, H. & Mevarech, M. (1993) *Biochemistry* 32, 4308-4313

Charlebois, R.L., Hofman, J.D., Schalkwyk, L.C., Lam, W.L. & Doolittle, W.F. (1990) *Can. J. Microbiol.* 35, 21-29

Charlebois, R.L., Lam, W.L., Cline, S.W. & Doolittle, W.F. (1987) *Proc. Natl. Acad. Sci. USA* 84, 8530-8534

Charlebois, R.L., Schalkwyk, L.C., Hofman, J.D. & Doolittle, W.F. (1991) *J. Mol. Biol.* 222, 509-524

Conover, R.K. & Doolittle, W.F. (1990) *J. Bacteriol.* 172, 3244-3249

Daniels, D.L., Plunkett, G., Burland, V. & Blattner, F.R. (1992) *Science* 257, 771-778

Danson, M.J. (1988) *Adv. Micro.Physiol.* 29, 165-231

Danson, M.J., Black, S.C., Woodland, D.L. & Wood, P.A. (1985) *FEBS Lett.* 179, 120-124

Danson, M.J., McQuattie, A. & Stevenson, K.J. (1986) *Biochemistry* 25, 3880-3884

David, M., Lubinsky-Mink, S., Ben-Zvi, A., Suissa, M., Ulitzur, S. & Kuhn, J. (1991) *Biochem. J.* 278, 225-234

Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nuc. Acids Res.* 12, 387-395

Distler, J., Mansouri, K., Mayer, G., Stockmann, M. & Piepersberg, W. (1992) *Gene* 115, 105-111

Distler, J., Ebert, A., Mansouri, K., Pissowotski, K., Stockmann, M. & Piepersberg, W. (1987) *Nuc. Acids Res.* 15, 8041-8056

Donald, L.J., Molgat,G.F. & Duckworth, H.W. (1989) *J. Bacteriol.* 171, 5542-5550

Donald, L.J. & Duckworth, H.W. (1987) *Biochem. Cell Biol.* 65, 930-938

Eisenberg, H., Mevarech M. & Zaccai, G. (1992) *Adv. Protein Chem.*, 43, 1-63

Evans, C.T., Owens, D.D., Sumegi, B., Kispal, G. & Srere, P.A. (1988) *Biochemistry* 27, 4680-4686

Forterre, P., Benachenhou-Lahfa, N., Confalonieri, M.D., Elie, C. & Labedan, B. (1993) *BioSystems* 28, 15-32

Fukaya, M., Takemura, H., Okumura, H., Kawamura, Y., Horinouchi, S. & Beppu, T. (1990) *J. Bacteriol.* 172, 2096-2104

Glaser, P., Kunst, F., Arnaud, M., Coudart, M.P., Gonzales, W., Hullo, M.F., Ionescu, M., Lubochinsky, B., Marcelino, L., Moszer, I., Presecan, E., Santana, M., Schneider, E., Schweitzer, J., Vertes, A., Rapoport, G. & Danchin, A. (1993) *Mol. Microbiol.* 10, 371-384

Goethals, K.H.M., Leyman, B., Van den Eeede, G., Van Montagu, M. (1993) Unpublished.

Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., Manolson, M.F., Poole, R.J., Date, T., Oshima, T., Konishi, J., Denda, K. & Yoshida, M. (1989) *Proc. Natl. Acad. Sci. USA* 86, 6661-6665

Gupta, M.N. (1992) *Eur. J. Biochem.* 203, 25-32

Gupta, R., Lanter, J.M. & Woese, C.R. (1983) *Science* 221, 656-659

Hammerschmidt, S., Birkholz, C., Robertson, B.D., van Putten, J., Ebeling, O. & Frosch, M. (1992) Unpublished

Heinzen, R.A., Frazier, M.E. & Mallavia, L.P. (1991) *Gene* 109, 63-69

Henneke, C.M., Danson, M.J., Hough, D.W. & Osguthorpe, D.J. (1989) *Protein Eng.* 2, 597-604

Heinzen, R.A., Frazier, M.E. & Mallavia, L.P. (1991) *Gene* 109, 63-69

Hickson, I (1994) Final year project report, Department of Biochemistry, University of Bath

Higa, H. & Cazzulo, J.J. (1975) *Biochem J.* 147, 267-274

Holmes, M.H., Nuttall, S.D. & Dyall-Smith, M.L. (1991) *J. Bacteriol.* 173, 3807-3813

Holmes, M.L., Dyall-Smith,M.L. (1991) J.Bacteriol. 173, 624-648

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. (1989) *Proc. Natl. Acad. Sci. USA* 86, 9355-9359

James, K.D. (1990) Final Year Project Report No.461, Department of Biochemistry, University of Bath

Jiang, X.M., Neal, B., Santiago, F., Lee S.J., Romana, L.K. & Reeves, P.R. (1991) *Mol. Microbiol.* 5, 695-713

Kandler, O. & Konig, H. (1994) In *The Biochemistry of the Archaea (Archaebacteria)*, 223-255. (Kates *et al.*, eds.) Elsevier Science Publishers B.V.

Karpusas, M., Branchaud, B. & Remington, S.J. (1990) *Biochemistry* 29, 2213-2219

Karpusas, M., Holland, D. & Remington, S.J. (1991) *Biochemistry* 30, 6024-6031

Kates, M. (1994) In *The Biochemistry of the Archaea (Archaebacteria)*, 261-295. (Kates *et al.*, eds.) Elsevier Science Publishers B.V.

Kirchberger, J., Cadelis, F., Kopperschläger, G. & Vijayalakshmi, M.A., Klibanov, A.M. (1989) *Trends Biochem. Sci.* 14, 141-144

Koeplin, R., Wang., Hoette, B., Priefer, U.B. & Puhler, A. (1993) Unpublished.

Kornfeld, S. & Glaser, L. (1961) *J. Biol. Chem.* 236, 1791

Lake, J.A. (1988) *Nature* 331, 184-186

Lake, J.A. (1989) *Can. J. Microbiol.* 35, 109-118

Lam, W.L. & Doolittle, W.F. (1989) *Proc. Natl. Acad. Sci. USA* 86, 5478-5482

Lam, W.L., Cohen, A., Tsouluhas, D. & Doolittle, W.F. (1990) *Proc. Natl. Acad. Sci. USA.* 87, 6614-6618

Lanyi, J.K. (1974) *Bacteriol. Rev.* 38 No.3, 272-290

Liao, D., Karpusas, M. & Remington S.J. (1991) *Biochemistry* 30, 6031-6036

Lill, U., Lefrank, S., Henschen, A. & Eggerer, H. (1992) *Eur. J. Biochem.* 208, 459-466

Lindquist, L., Kaiser, R., Reeves, P.R. & Lindberg, A.A. (1993) *Eur. J. Biochem.* 211, 763-770

Man, W., Li, Y., O'Connor, C.D. & Wilton, D.C. (1991) *Biochem. J.* 280, 521-526

Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory

Margulis, L. & Guerrero, R. (1991) *New Scientist* 23rd March, 46-50

Mayr, E. (1990) *Nature* 348, 491

Mayr, E. (1991) *Nature* 353, 122

McCabe, P.C., in *PCR Protocols: A Guide to Methods and Applications* (1990) Eds. Innis, M.A., Gelfland, D.H., Sninsky, J.J. & White, T.J.

Mevarech, M. & Werczberger, R. (1985) *J. Bacteriol.* 162, 461-462

Mitchell, C.G. & Weitzman, P.D.J. (1983) *FEBS Lett.* 151 No.2, 260-264

Muir, J.M., Hough, D.W. & Danson, M.J. (1994)a *System. Appl. Microbiol.* 16, 528-533

Muir, J.M., Hough, D.W. & Danson, M.J. (1994)b Unpublished data.

Mullakhanbhai, M.F. & Larsen, H. (1975) *Arch. Microbiol.* 104, 207-214

Ner, S.S., Bhayana, V., Bell, A.W., Gile, I.G., Duckworth, H.W. & Bloxham, D.P. (1983) Biochemistry 22, 5243-5249

Numata, O., Takemasa, T., Takagi, I., Hirono, M., Hirano, H., Chiba, J. & Watanabe, Y. (1991) *Biochem. Biophys. Res. Comm.* 174, 1028-1034

Okazaki, R., Okazaki, T., Strominger, J.L. & Michelson, A.M. (1962) *J. Biol. Chem.* 237, 3014-3026

Olsen, G.J. & Woese, C.R. (1989) *Can. J. Microbiol.* 35, 119-122

Olsen, G.J. & Woese, C.R. (1993) *FASEB J.* 7, 113-123

Parker, J.M.R., Gou, D. & Hodges, R.S. (1986) *Biochemistry* 258, 5425-5432

Puehler, G., Lottspeich, F & Zillig, W. (1989) *Nucleic Acid Res.* 17, 4517-4534

Rajakumar, K., Jost, H., Sasakawa, C., Okada, N., Yoshikawa, M. & Adler, B. (1994) Unpublished.

Rao, M.J.K. & Argos, P. (1981) *Biochemistry* 20, 6536-6543

Reiter, W.D., Hudepohl, U. & Zillig, W. (1990) *Proc. Natl. Acad. Sci. USA* 87, 9509-9513

Remington, S.J. (1992) *Curr. Top. Cell. Regul.* 33, 209-229

Remington, S.J., Wiegand, G. & Huber, R. (1982) *J. Mol. Biol.* 158, 111-152

Richardson, J.S. & Richardson, D.C. (1988) *Science* 240 1648-1652

Rodriguez-Valera, F. (1991) *Biochem. Soc. Symp.* 58, 135-147

Rodriguez-Valera, F., Lillo, J.A.G., Antón, J. & Meseguer, I. (1989) *General and Applied Aspects of Halophilic Microorganisms*, 373-380 (F. Rodriguez-Valera, ed.), Plenum Press, New York.

Rosenkrantz, M., Alam, T., Kim, K., Clark, B.J., Srere, P.A. & Guarente, L.P. (1986) *Mol. and Cell. Biol.* 6, 4509-4515

Russell, R.J.M., Byrom, D., Danson, M.J.D., Hough, D.W. & Taylor, G.L. (1993) *J. Mol. Biol.* 232, 308-309

Ryu, K. & Dordick, J.S. (1993) *J. Amer. Chem. Soc.* 205, 9 (Abstract)

Schäfer, S., Barkowski, C. & Fuchs, G. (1986) *Arch. Microbiol.* 146, 301-308

Schendel, F.J., August, P.R., Anderson, C.R., Hanson, R.S. & Flickinger, M.C. (1992) *Appl. and Environ. Microbiol.* 58, 335-345

Schwartz, R.M. & Dayhoff, M.O. (1979) in *Atlas of Protein Sequence and Structure* (Dayhoff, M.O., ed.), National Biomedical Research Foundation, Washington D.C., 353-358

Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517

Srere, P.A., Brazil, H. & Gonen, L. (1963) *Acta Chem. Scand.* 17, S129-S134

Sutherland, K.J., Henneke, C.M., Towner, P., Hough, D.W. & Danson, M.J. (1990) Eur. J. Biochem. 194, 839-844

Tautz, D. & Renz, M. (1983) *Anal. Biochem.* 132, 14-19

Vettakkorumakankav, N., Danson, M.J. Hough, D.W., Stevenson, K.J., Davison, M. & Young, J. (1992) *Biochem. Cell Biol.* 70, 70-75

Wang, Z., Srere, P.A. & Evans, C.T. (1991) *Biochemistry* 30, 9281-9286

Wang, L., Romana, L.K. & Reeves, P.R. (1992) *Genetics* 103, 429-443

Weitzman, P.D.J. & Ridley, J. (1983) *Biochem. and Biophys. Res. Comm.* 112, 1021-1026

Wheelis, M.L., Kandler, O. & Woese, C.R. (1992) *Proc. Natl. Acad. Sci. USA* 89, 2930-2934

Woese, C.R. & Fox, G.E. (1977) *Proc. Natl. Acad. Sci. USA* 74 No.11, 5088-5090

Woese, C.R., Kandler, O. & Wheelis, M.L. (1990) *Proc. Natl. Acad. Sci. USA* 87, 4576-4579

Woese, C.R., Kandler, O. & Wheelis, M.L. (1991) *Nature* 351, 528

Woese, C.R., Magrum, L.J & Fox, G.E. (1978) *J. Mol. Evol.* 11, 245-252

Wood, D.O., Williamson, L.R., Winkler, H.H. & Krause, D.C. (1987) *J. Bacteriol.* 169, 3564-3572

Zaccai, G & Cendrin, F (1989) *J. Mol. Biol.* 208, 491-500

Zaccai, G & Eisenberg, H. (1990) *Trends Biochem. Sci.* 15, 333-337

Zhang, L., Al-Hendy, A., Toivanen, P. & Skurnik, M. (1993) *Mol. Microbiol.* 9, 309-321

Zillig, W., Klenk, H., Palm, P., Leffers, H., Pühler, G., Gropp, F. & Garrett, R.A. (1989a) *Endocytobiosis & Cell Res.* 6, 1-25

Zillig, W., Klenk, H., Palm, P., Pühler, G., Gropp, F., Garrett, R.A. & Leffers, H. (1989b) *Can. J. Microbiol.* 35, 73-80

Zillig, W., Palm, P., Reiter, W.D., Gropp, F., Puhler, G. & Klenk, H. (1988) *Eur. J. Biochem.* 173, 473-482

Zusman, T., Rosenshine, I., Boehm, G., Jaenicke, R., Leskiw, B. &
Mevarech, M. (1989) *J. Biol. Chem.* 264 No.32, 18878-18883